

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editors: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

ARTHUR C. CARR, *New York State Psychiatric Institute and Columbia University*

CONSULTING EDITORS

- | | |
|--|--|
| THEODORE X. BARBER
<i>Medfield Research Foundation</i> | LAWRENCE C. KOLB
<i>New York State Psychiatric Institute
and Columbia University</i> |
| HAROLD BASOWITZ
<i>New York University</i> | SHELDON J. KORCHIN
<i>University of California</i> |
| RICHARD L. BLUMENTHAL
<i>New York State Psychiatric Institute
and Columbia University</i> | WILLIAM G. LAWLOR
<i>Fordham University</i> |
| EUGENE I. BURDOCK
<i>New York State University</i> | ARDIE LUBIN
<i>United States Naval Hospital, San Diego</i> |
| JOHN M. BUTLER
<i>University of Chicago</i> | SALVATORE MADDI
<i>University of Chicago</i> |
| LOREN J. CHAPMAN
<i>University of Wisconsin</i> | MARTIN T. ORNE
<i>Institute of the Pennsylvania Hospital
and the University of Pennsylvania</i> |
| W. CRAWFORD CLARK
<i>New York State Psychiatric Institute
and Columbia University</i> | ERNST PRELINGER
<i>Yale University</i> |
| JOHN H. CUMMING
<i>New York State Department of Mental
Hygiene</i> | DAVID ROSENTHAL
<i>National Institute of Mental Health</i> |
| W. GRANT DAHLSTROM
<i>University of North Carolina</i> | ALAN K. ROSENWALD
<i>Illinois State Psychiatric Institute</i> |
| JOSEPH L. FLEISS
<i>New York State Department of Mental
Hygiene and Columbia University</i> | JACQUES RUTSCHMANN
<i>New York State Psychiatric Institute and
Columbia University</i> |
| JEROME D. FRANK
<i>Johns Hopkins University</i> | KURT SALZINGER
<i>New York State Psychiatric Institute and
Brooklyn Polytechnic Institute</i> |
| SOL L. GARFIELD
<i>Columbia University</i> | JOSEPH SCHACHTER
<i>New York State Psychiatric Institute
and Columbia University</i> |
| ANN MACARET GARNER
<i>University of Nebraska</i> | JOHN SHLIEN
<i>University of Chicago</i> |
| ISRAEL GOLDIAMOND
<i>Institute for Behavioral Research</i> | MORRIS I. STEIN
<i>New York University</i> |
| HARRISON G. GOUGH
<i>University of California, Berkeley</i> | HAROLD L. WILLIAMS
<i>University of Oklahoma School of Medicine</i> |
| STARKE R. HATHAWAY
<i>University of Minnesota</i> | JOSEPH ZUBIN
<i>New York State Psychiatric Institute
and Columbia University</i> |
| RALPH HEINE
<i>University of Michigan</i> | |
| ERNEST R. HILGARD
<i>Stanford University</i> | |
| HELEN ORR
Managing Editor | MYRTLE HENDRICKS
Technical Editor |
| | HENRY S. SALMON
Advertising and
Business Manager |
| | ROBERT D. GORMAN
Subscription Manager |

VOLUME 74
1969

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

Prince and Lemon Sts., Lancaster Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036
Copyright © 1969 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 74

ABRAMOFF, EILEEN. See ENDICOTT, NOBLE A.	
ADINOLFI, ALLEN A. See GOLDFRIED, MARVIN R.	
AGRAS, W. STEWART. See BARLOW, DAVID H.	
AGRAS, W. STEWART. See LEITENBERG, HAROLD.	
ANDERSON, V. ELVING, SIEGEL, FELICIA S., FISCH, ROBERT O., & WIRT, ROBERT D. Responses of Phenylketonuric Children on a Continuous Performance Test	358
ARONSON, H., & WEINTRAUB, WALTER. Certain Initial Variables as Predictors of Change with Classical Psychoanalysis	490
ARTHUR, RANSOM J. See GUNDERSON, E. K. ERIC.	
BAER, PAUL E., & FUHRER, MARCUS J. Cognitive Factors in Differential Conditioning of the GSR: Use of a Reaction Time Task as the UCS with Normals and Schizophrenics	544
BAKER, BRUCE L. Symptom Treatment and Symptom Substitution in Enuresis	42
BARBER, THEODORE XENOPHON, & CALVERLEY, DAVID SMITH. Multidimensional Analysis of "Hypnotic" Behavior	209
BARLOW, DAVID H. See LEITENBERG, HAROLD.	
BARLOW, DAVID H., LEITENBERG, HAROLD, & AGRAS, W. STEWART. Experimental Control of Sexual Deviation through Manipulation of the Noxious Scene in Covert Sensitization	596
BARRELL, ROBERT P. See DEWOLFE, ALAN S.	
BARRETT, CURTIS L. Systematic Desensitization versus Implosive Therapy	587
BAUMEISTER, ALFRED A., & WILCOX, STEPHEN J. Effects of Variations in the Preparatory Interval on the Reaction Times of Retardates and Normals	438
BAYER, L. See ZUBEK, JOHN P.	
BEAKER, NANCY G., & MEHRABIAN, ALBERT. Inconsistent Communications and Psychopathology	126
BECKER, JOSEPH, & FINKEL, PAUL. Predictability and Anxiety in Speech by Parents of Female Schizophrenics	517
BECKER, JOSEPH, & IWAKAMI, EILEEN. Conflict and Dominance within Families of Disturbed Children	330
BECKER, JOSEPH, & STEPKES, HERBERT. Parental Dominance, Conflict, and Disciplinary Coerciveness in Families of Female Schizophrenics	193
BEHAR, LENORE, & SPENCER, ROGER F. Relationship between Psychosocial Adjustment and Perception of Maternal Attitudes	471
BENJAMIN, THOMAS B., & WATT, NORMAN F. Psychopathology and Semantic Interpretation of Ambiguous Words	706
BENTLER, P. M., & PRINCE, CHARLES. Personality Characteristics of Male Transvestites: III	140
BERGIN, ALLEN E., & JASPER, LAWRENCE G. Correlates of Empathy in Psychotherapy: A Replication	477
BERTRAND, SHARON, & MASLING, JOSEPH. Oral Imagery and Alcoholism	50
BERZINS, JURIS I., FRIEDMAN, WILLIAM H., & SEIDMAN, EDWARD. Relationship of the A-B Variable to Patient Symptomatology and Psychotherapy Expectancies	119
BIGGS, BARBARA, & SHEEHAN, JOSEPH. Punishment of Distraction? Operant Stuttering Revised	256
BLANTON, RICHARD L. See DECKNER, C. WILLIAM.	
BLOCH, ANNE-MARIE ALLERAND. Remembrance of Feelings Past: A study of Phenomenological Genetics	340
BOSTWICK, SUSAN. See JEROME, EDWARD A.	
BOURNE, LYLE E., JR. See PISHKIN, VLADIMIR.	
BOWERS, KENNETH S., & GILMORE, J. BARNARD. Subjective Report and Credibility: An Inquiry Involving Hypnotic Hallucinations	443
BRELJE, TERRY. See CHAPMAN, LOREN J.	
BRIDGER, WAGNER H. See COOPER, ALLAN.	
BRONSON, LOUISE. See MEADOW, ARNOLD.	
BUCHWALD, ALEXANDER M. See FRANKEL, A. STEVEN.	
CALMAS, WILFRED. See COHEN, MURRAY.	
CALVERLEY, DAVID SMITH. See BARBER, THEODORE XENOPHON.	
CANCRO, ROBERT, & SUGERMAN, A. ARTHUR. Psychological Differentiation and Process-Reactive Schizophrenic	415
CHAPMAN, JEAN P. See CHAPMAN, LOREN J.	
CHAPMAN, LOREN J., & CHAPMAN, JEAN P. Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs	271
CHAPMAN, LOREN J., CHAPMAN, JEAN P., & BRELJE, TERRY. Influence of the Experimenter on Pupillary Dilation to Sexual Provocative Pictures	396
CHAPMAN, LOREN J. See KLORMAN, RAFAEL.	
CLARK, W. CRAWFORD. Sensory-Decision Theory Analysis of the Placebo Effect on the Criterion for Pain and Thermal Sensitivity (d')	363
CLEVELAND, SIDNEY E. See FISHER, SEYMOUR.	
COHEN, BERTRAM D. See NACHMANI, GILEAD.	

COHEN, MURRAY, SEGHERN, THEOHARIS, & CALMAS, WILFRED. Sociometric Study of the Sex Offender	249
COLE, C. W., OETTING, E. R., & MISKIMINS, R. W. Self-Concept Therapy for Adolescent Females	642
CONDON, WILLIAM S., OGSTON, WILLIAM D., & PACOE, LARRY V. Three Faces of Eve Revisited: A Study of Transient Microstrabismus	618
COOPER, ALLAN, FURST, JOSEPH B., & BRIDGER, WAGNER H. A Brief Commentary on the Usefulness of Studying Fears of Snakes	413
CROMWELL, RUE L. See NEALE, JOHN M.	
DAVISON, GERALD C. A Procedural Critique of "Desensitization and the Experimental Reduction of Threat"	86
DECKNER, C. WILLIAM, & BLANTON, RICHARD L. Effect of Context and Strength of Association on Schizophrenic Verbal Report	348
DEL CASTILLO, DAVID. See KOEING, KARL P.	
DRENNEN, WILLIAM, GALLMAN, WILLIAM, & SAUSSER, GENE. Verbal Operant Conditioning of Hospitalized Psychiatric Patients	454
DENNERLL, RAYMOND D. See SCHWARTZ, MELVIN L.	
DEWOLFE, ALAN S., BARRELL, ROBERT P., & SPANER, FRED E. Staff Attitudes toward Patient Care and Treatment-Disposition Behavior	90
DEWOLFE, ALAN S. See FORGUS, RONALD.	
DIES, ROBERT R. See FARINA, AMERIGO.	
DOLBY, LISE-LOTTE LEHD. See HIGGINS, JERRY.	
DOUGLAS, VIRGINIA I. See FREIDBERGS, VAIRA.	
DUFORT, W. HENRY. See STRAUGHAN, JAMES H.	
DWORKIN, EARL. See GOLDFRIED, MARVIN R.	
EFRAIM, JAY S. See MARCIA, JAMES E.	
ENDICOTT, NOBLE A., JORTNER, SIDNEY, & ABRAMOFF, EILEEN. Objective Measures of Suspiciousness	26
ERIKSON, ROBERT V. See ROBERTS, ALAN H.	
ETAUGH, CLAIRE FALK. See HURT, MICHAEL L.	
EVANS, KAREN LAWSON. See FOLKINS, CARLYLE H.	
FARINA, AMERIGO, HOLZBERG, JULES D., & DIES, ROBERT R. Influence of the Parents and Verbal Reinforcement on the Performance of Schizophrenic Patients	9
FEFFER, MELVIN. See LOWENHERZ, LILA.	
FERREIRA, ANTONIO J. See WINTER, WILLIAM D.	
FINKEL, PAUL. See BECKER, JOSEPH.	
FISCH, ROBERT O. See ANDERSON, V. ELVING.	
FISHER, SEYMOUR, & CLEVELAND, SIDNEY E. Rejoinder to Hirt and Kurtz' "A re-examination of the Relationship between Body Boundary and Site of Disease" ..	144
FOLKINS, CARLYLE H., EVANS, KAREN LAWSON, OPTON, EDWARD M., JR., & LAZARUS, RICHARD S. A Reply to Davison's Critique	88
FONG, STANLEY L., & PESKIN, HARVEY. Sex-Role Strain and Personality Adjustment of China-Born Students in America: A Pilot Study	563
FORGUS, RONALD, & DEWOLFE, ALAN S. Perceptual Selectivity in Hallucinatory Schizophrenics	288
FOX, ROBERT. See NEALE, JOHN M.	
FRANKEL A. STEVENS, & BUCHWALD, ALEXANDER M. Verbal Conditioning of Common Associations in Long-Term Schizophrenics: A Failure	372
FREIBERGS, VAIRA, & DOUGLAS, VIRGINIA I. Concept Learning in Hyperactive and Normal Children	388
FRIEDMAN, WILLIAM H. See BERZINS, JURIS I.	
FUHRER, MARCUS J. See BAER, PAUL E.	
FURST, JOSEPH B. See COOPER, ALLAN.	
GALLMAN, WILLIAM. See DRENNEN, WILLIAM.	
GASSNER, SUZANNE, & MURRAY, EDWARD J. Dominance and Conflict in the Interactions between Parents of Normal and Neurotic Children	33
GEER, JAMES H., & KLEIN, KENNETH. Effects of Two Independent Stresses upon Automic Responding	237
GILMORE, J. BARNARD. See BOWERS, KENNETH S.	
GLESER, GOLDINE C. See HIRT, MICHAEL.	
GOLDFRIED, MARVIN R., ADINOLFI, ALLEN A., & DWORKIN, EARL. Effect of Developmentally Relevant Information and Severity of Disturbance on the Understanding and Evaluation of Deviant Behavior	482
GOLDMAN, RUTH K., MENDELSON, GERALD A. Psychotherapeutic Change and Social Adjustment: A Report of a National Survey of Psychotherapists	164
GOLDSTEIN, STEVEN G., & LINDEN, JAMES D. Multivariate Classification of Alcoholics by Means of the MMPI	661
GOSS, ALLEN, & MOROSKO, THOMAS E. Alcoholism and Clinical Symptoms	682
GOYEN, JUDITH. See LYLE, J. G.	
GUNDERSON, E. K. ERIC, & ARTHUR, RANSOM J. A Brief Mental Health Index	100

HARFORD, THOMAS, & SOLOMON, LEONARD. Effects of "Reformed Sinner" and a "Lapsed Saint" Strategy upon Trust Formation in Paranoid and Nonparanoid Schizophrenic Patients	498
HARRIS, MARY B. Self-Directed Program for Weight Control: A Pilot Study	263
HARRISON, ROBERT H. See ROSSI, A. MICHAEL.	
HEINICKE, CHRISTOPH M. Frequency of Psychotherapeutic Session as a Factor Affecting Outcome: Analysis of Clinical Ratings and Test Results	553
HIGGINS, JERRY, PETERSON, JUDITH C., & DOLBY, LISE-LOTTE LEHD. Social Adjustment and Familial Schema	296
HIRT, MICHAEL, & KURTZ, RICHARD. A Reexamination of the Relationship between Body Boundary and Site of Disease	67
HIRT, MICHAEL, ROSS, W. DONALD, KURTZ, RICHARD, & GLESSER, GOLDINE C. Attitudes to Body Products among Normal Subjects	486
HOLDEN, EDWARD A., JR. See THOR, DONALD.	
HOLZBERG, JULES D. See FARINA, AMERIGO.	
HURT, MICHAEL L., & ETAUGH, CLAIRE FALK. Effects of Conditioning upon Preexperimentally Acquired Verbal Habit in Psychiatric Patients	583
IRWIN, LOUISE, & RENNER, K. EDWARD. Effect of Praise and Censure on the Performance of Schizophrenics	221
JASPER, LAWRENCE G. See BERGIN, ALLEN E.	
JEROME, EDWARD A., YOUNG, MARGUERITE L., & BOSTWICK, SUSAN. Reality Testing by Schizophrenics	148
JORTNER, SIDNEY. See ENDICOTT, NOBLE A.	
KAPCHE, ROBERT. Associative Response Competition in Schizophrenia	75
KATES, SOLIS L. Learning and Use of Logical Symbols by Deaf and Hearing Subjects	699
KATKIN, EDWARD S. & MCCUBBIN, ROBERT J. Habituation of the Orienting Response as a Function of Individual Differences in Anxiety and Autonomic Lability	54
KATKIN, EDWARD S. See STARR, B. JAMES.	
KATZ, PHYLLIS A., & ROSENBERG, SUE. Effects of Labels on the Perception and Discrimination Learning of Retardates	95
KEMPLER, BERNHARD. See O'NEIL, MARION.	
KIRCHNER, ELIZABETH P. See PIERS, ELLEN V.	
KLEIN, DONALD F. See ROSEN, BERNARD.	
KLEIN, KENNETH. See GEER, JAMES H.	
KLETT, C. JAMES. See LORR, MAURICE.	
KLOORMAN, RAFAEL, & CHAPMAN, LOREN J. Regression in Schizophrenic Thought Disorder	199
KOEING, KARL P., & DEL CASTILLO, DAVID. False Feedback and Longevity of the Conditioned GSR during Extinction: Some Implication for Aversion Therapy	505
KRAUSE, MERTON S. Construct Validity for the Evaluation of Therapy Outcomes	524
KUN, KAROLY J. See VOGEL, WILLIAM.	
KURTZ, RICHARD. See HIRT, MICHAEL.	
LANG, PETER J., & MELAMED, BARBARA G. Case Report: Avoidance Conditioning Therapy of an Infant with Chronic Ruminative Vomiting	1
LANGER, JONAS, STEIN, KENNETH B., & ROSENBERG, B. G. Cognitive Interference by Nonverbal Symbols in Schizophrenics	474
LAZARUS, RICHARD S. See FOLKINS, CARLYLE H.	
LEIBOWITZ, HERSCHEL W. See PARRISH, MICHAEL.	
LEITENBERG, HAROLD, AGRAS, W. STEWART, BARLOW, DAVID H., & OLIVEAU, DONALD C. Contribution of Selective Positive Reinforcement and Therapeutic Instructions to Systematic Desensitization Therapy	113
LEITENBERG, HAROLD. See BARLOW, DAVID H.	
LEVENSTEIN, SIDNEY. See ROSEN, BERNARD.	
LIN, YI-GUANG. See SCHWARTZ, MELVIN L.	
LINDEN, JAMES D. See GOLDSTEIN, STEVEN G.	
LITVAK, STUART B. See RIMM, DAVID C.	
LOFTIN, JULIE. See SCHOPLER, ERIC.	
LORR, MAURICE, & KLETT, C. JAMES. Cross-Cultural Comparison of Psychotic Syndromes	531
LOWENHERZ, LILA, & FEFFER, MELVIN. Cognitive Level as a Function of Defensive Isolation	352
LUNDY, RICHARD. See PARRISH, MICHAEL.	
LYLE, J. G., & GOYEN, JUDITH. Performance of Retarded Readers on the WISC and Educational Tests	105
MACK, JAMES L. The MMPI and Recidivism	612
MAGARO, PETER A. Size Estimation in Schizophrenia as a Function of Censure Diagnosis, Premorbid Adjustment, and Chronicity	306
MARCIA, JAMES E., RUBIN, BARRY M., & EFRAN, JAY, S. Systematic Desensitization: Expectancy Change or Counterconditioning?	382
MARKEL, NORMAN N. Relationship between Voice-Quality Profiles and MMPI Profiles in Psychiatric Patients	61

MARTIN, DAVID G. Consistency of Self-Descriptions under Different Role Sets in Neurotic and Normal Adolescents and Adults	173
MASLING, JOSEPH. See BERTRAND, SHARON.	
MAYMAN, MARTIN, & VOTH, HAROLD M. Reality Closeness, Phantasy, and Autokinesis: A Dimension of Cognitive Style	635
McCUBBIN, ROBERT J. See KATKIN, EDWARD S.	
McINTYRE, CURTIS W. See NEALE, JOHN M.	
McKINNON, G., & SINGER, G. Schizophrenia and the Scanning Cognitive Control: A Reevaluation	242
McREYNOLDS, WILLIAM T. A Note on Relaxation Treatment Groups in Studies of Systematic Desensitization	561
MEADOW, ARNOLD, & BRONSON, LOUISE. Religious Affiliation and Psychopathology in Mexican-American Population	177
MEHRABIAN, ALBERT. See BEAKEI, NANCY G.	
MELAMED, BARBARA G. See LANG, PETER J.	
MENDELSON, GERALD A. See GOLDMANN, RUTH K.	
MENDELSON, GERALD A., & RANKIN, NEIL O. Client-Counselor Compatibility and the Outcome of Counseling	157
MESHORER, EDWARD. See KUN, KAROLY J.	
MILSTEIN, S. See ZUBEK, JOHN P.	
MINARD, JAMES G., & MOONEY, WILLIAM. Psychological Differentiation and Perceptual Defense: Studies of the Separation of Perception from Emotion	131
MINTZ, SANFORD. Effect of Actual Stress on Word Associations	293
MISKIMINS, R. W. See COLE, C. W.	
MITCHELL, KEVIN M. Concept of "Pathogenesis" in Parents of Schizophrenic and Normal Children	423
MOONEY, WILLIAM. See MINARD, JAMES G.	
MOOS, RUDOLF H. Sources of Variance in Responses to Questionnaires and in Behavior	405
MORISHIGE, HOWARD. See REYHER, JOSEPH.	
MOROSKO, THOMAS E. See ALLEN, GOSS.	
MURRAY, E. NEIL. Anxiety and Overgeneralization: Negative Results	602
MURRAY, EDWARD J. See GASSNER, SUZANNE.	
NACHMAN, GILEAD, & COHEN, BERTRAM D. Recall and Recognition Free Learning in Schizophrenics	511
NASH, MICHAEL M., & ZIMRING, FRED M. Prediction of Reaction to Placebo	568
NATHAN, PETER E. See ROSSI, A. MICHAEL.	
NEALE, JOHN M., McINTYRE, CURTIS W., FOX, ROBERT, & CROMWELL, RUE L. Span of Apprehension in Acute Schizophrenics	593
OETTING, E. R. See COLE, C. W.	
OGSTON, WILLIAM D. See CONDON, WILLIAM S.	
OLIVEAU, DONALD C. See LEITENBERG, HAROLD.	
O'NEIL, MARION, & KEMPLER, BERNHARD. Approach and Avoidance Responses of the Hysterical Personality to Sexual Stimuli	300
OPTON, EDWARD M., JR. See FOLKINS, CARLYLE H.	
ORRIS, JAMES BURDEANE. Visual Monitoring Performance in Three Subgroups of Male Delinquents	227
PACOE, LARRY V. See CONDON, WILLIAM S.	
PARRISH, MICHAEL, LUNDY, RICHARD M., & LEIBOWITZ, HERSCHEL W. Effect of Hypnotic Age Regression on the Magnitude of the Ponzo and Poggerdorff Illusions ..	693
PAUL, GORDON L. Physiological Effects of Relaxation Training and Hypnotic Suggestion	425
PESKIN, HARVEY. See FONG, STANLEY L.	
PETERSON, JUDITH C. See HIGGINS, JERRY.	
PIERS, ELLEN V., & KIRCHNER, ELIZABETH P. Eyelid Conditioning and Personality: Positive Results from Nonpartisans	336
PISHKIN, VLADIMIR, & BOURNE, LYLE E., JR. Concept Identification by Schizophrenic and Normal Subjects as a Function of Problem Complexity Relevance of Social Cues	314
PRINCE, CHARLES. See BENTLER, P. M.	
RANKIN, NEIL O. See MENDELSON, GERALD A.	
RENNER, K. EDWARD. See IRWIN, LOUISE.	
REYHER, JOSEPH. Comment on "Artificial Induction of Posthypnotic Conflict"	420
REYHER, JOSEPH, & MORISHIGE, HOWARD. Electroencephalogram and Rapid Eye Movements during Free Imagery and Dream Recall	576
REYHER, JOSEPH. See SANDERS, RAYMOND S. JR.	
RIMM, DAVID C., & LITVAK, STUART B. Self-Verbalization and Emotional Arousal	181
ROBERTS, ALAN H., & ERIKSON, ROBERT V. Measuring Impulse Control in Institutionalized Delinquents Using Rorschach Content and Thought Process Scales ..	632
ROSEN, BERNARD, KLEIN, DONALD F., LEVENSTEIN, SIDNEY, & SHAHINIAN, SIROON P. Social Competence and Posthospital Outcome among Schizophrenic and Non-schizophrenic Psychiatric Patients	401
ROSENBERG, B. G. See LANGER, JONAS.	

ROSENBERG, SUE. See KATZ, PHYLLIS A.	
ROSS, W. DONALD. See HIRT, MICHAEL.	
ROSSI, A. MICHAEL, NATHAN, PETER E., HARRISON, ROBERT H., & SOLOMON, PHILIP. Operant Responding for Visual Stimuli during Sensory Deprivation: Effect of Meaningfulness	188
RUBIN, BARRY M. See MARCIA, JAMES E.	
SANDERS, RAYMOND S., JR., & REYHER, JOSEPH. Sensory Deprivation and the Enhancement of Hypnotic Susceptibility	375
SAUSSER, GENE. See DRENNEN, WILLIAM.	
SCHOOLER, CARMI, & SILVERMAN, JULIAN. Perceptual Styles and Their Correlates among Schizophrenic Patients	459
SCHOPLER, ERIC, & LOFTIN, JULIE. Thinking Disorders in Parents of Young Psychotic Children	281
SCHWARTZ, MELVIN L., DENNERLL, RAYMOND D., & LIN, YI-GUANG. Similarity of Personality Trait Interrelationships in Persons with and without Epileptogenic Cerebral Dysfunction	205
SEGHORN, THEOHARIS. See COHEN, MURRAY.	
SEIDMAN, EDWARD. See BERZINS, JURIS I.	
SHAHINIAN, SIROON P. See ROSEN, BERNARD.	
SHEEHAN, JOSEPH. Cyclic Variation in Stuttering: Comment on Taylor and Taylor's "Test of Predictions from the Conflict Hypothesis of Stuttering"	452
SHEEHAN, JOSEPH. See BIGGS, BARBARA.	
SHEEHAN, PETER W. Artificial Induction of Posthypnotic Conflict	16
SHEPHARD, JEAN MARY. See ZUBEK, JOHN P.	
SIEPKES, HERBERT. See BECKER, JOSEPH.	
SIEGEL, FELICIA, S. See ANDERSON, V. ELVING.	
SILVERMAN, JULIAN. See SCHOOLER, CARMI.	
SINGER, G. See MCKINNON, T.	
SKRZYPEK, GEORGE J. Effect of Perceptual Isolation and Arousal on Anxiety, Complexity preference, and Novelty Preference in Psychopathic and Neurotic Delinquents	321
SOLOMON, LEONARD. See HARFORD, THOMAS.	
SOLOMON, PHILIP. See ROSSI, A. MICHAEL.	
SPANER, FRED E. See DEWOLFE, ALAN S.	
SPENCER, ROGER F. See BEHAR, LENORE.	
STARR, B. JAMES, & KATKIN, EDWARD S. The Clinician as an Aberrant Actuary: Illusory Correlation and the Incomplete Sentences Blank	670
STEELE, JAMES. The Hysteria and Psychasthenia Constructs as an Alternative to Manifest Anxiety and Conflict-Free Ego Functions	79
STEIN, KENNETH B. See LANGER, JONAS.	
STRAUGHAN, JAMES H., & DUFORT, W. HENRY. Task Difficulty, Relaxation, and Anxiety Level during Verbal Learning and Recall	621
STREINER, DAVID L. Effects of Task Complexity and Verbal Evaluation on the Learning of Normals and Schizophrenics	606
SUGERMAN, A. ARTHUR. See CANCRO, ROBERT.	
SULLIVAN, RICHARD. Subjective Matching of Anxiety to Intensities of White Noise ..	646
THOR, DONALD H., & HOLDEN, EDWARD A., JR. Visual Perception of Sequential Numerosity by Normals and Retardates	676
VOGEL, WILLIAM, KUN, KAROLY J., & MESHORER, EDWARD. Determinants of Institutional Release and Prognosis in Mental Retardates	685
VOTH, HAROLD M. See MAYMAN, MARTIN.	
WACHTEL, PAUL L. Psychology, Metapsychology, and Psychoanalysis	651
WATT, NORMAN F. See BENJAMIN, THOMAS B.	
WEINTRAUB, WALTER. See ARONSON, H.	
WILCOX, STEPHEN J. See BAUMEISTER, ALFRED A.	
WINTER, WILLIAM D., & FERREIRA, ANTONIO J. Talking Time as an Index of Intra-familial Similarity in Normal and Abnormal Families	574
WIRT, ROBERT D. See ANDERSON, V. ELVING.	
WOOD, FRANK A. An Investigation of Methods of Presenting Incomplete Sentence Stimuli	71
WRIGHT, NANCY A., & ZUBEK, JOHN P. Relationship between Perceptual Deprivation Tolerance and Adequacy of Defenses as Measured by the Rorschach	615
YOUNG, MARGUERITE L. See JEROME, EDWARD A.	
ZIMRING, FRED M. See NASH, MICHAEL M.	
ZUBEK, JOHN P., BAYER, L., MILSTEIN, S., & SHEPHARD, JEAN MARY. Behavioral and Physiological Changes during Prolonged Immobilization Plus Perceptual Deprivation	230
ZUBEK, JOHN P., BAYER, L., & SHEPHARD, JEAN MARY. Relative Effects of Prolonged Social Isolation and Confinement: Behavioral and EEG Changes	625
ZUBEK, JOHN P. See WRIGHT, NANCY A.	

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
*New York State Psychiatric Institute
and Columbia University*

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt

Journal of Abnormal Psychology

Department of Psychology

N. Y. State Psychiatric Institute & Columbia University

722 West 168th St.

New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR	MYRTLE HENDRICKS	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

Prince and Lemon Sts., Lancaster, Pa. 17604

and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

CASE REPORT:

AVOIDANCE CONDITIONING THERAPY OF AN INFANT WITH CHRONIC RUMINATIVE VOMITING¹

PETER J. LANG AND BARBARA G. MELAMED²

University of Wisconsin

This paper reports the treatment of a 9-mo.-old male infant whose life was seriously endangered by persistent vomiting and chronic rumination. An aversive conditioning paradigm, employing electric shock, significantly reduced the frequency of this maladaptive response pattern in a few, brief treatment sessions. Electromyographic records were used in assessing response characteristics of the emesis, and in determining the shock contingencies used in therapy. Cessation of vomiting and rumination was accompanied by weight gains, increased activity level, and general responsiveness to people.

A variety of techniques have been used in the treatment of persistent vomiting in infants and children. In general these therapies are tailored to the known or hypothesized causes of the disorder. Thus, the presence of functional disturbance in the intestinal tract would encourage the use of pharmacologic agents—"tranquilizers," antinauseants, or antiemetics. If gastric, anatomical anomalies can be diagnosed, their surgical removal often proves to be the most effective treatment. Animal studies suggest that surgical manipulation of the central nervous system may also become a vehicle for emesis control (Borison, 1959).

When diagnosis excludes obvious, organic antecedents, both the etiology and treatment of the disorder appear less certain. However,

clinical workers have described an apparently "psychosomatic" vomiting in children which is generally accompanied by a ruminative re-chewing of the vomitus. In reviewing the syndrome, Richmond, Eddy, and Green (1958) adhere to the widely held psychoanalytic hypothesis that it results from a disruption in the mother-infant relationship. They suggest that the condition is brought about by the inability of the mother to fulfill an adult psychosexual role which is reflected in marital inadequacy. She is unable to give up her own dependent needs and is incapable of providing warm, comfortable, and intimate physical care for the infant. This lack of comfort from without causes the infant to seek and recreate such gratification from within. Thus, in attempting to regain some satisfaction from the feeding situation, he regurgitates his food and retains it in his mouth. The recommended treatment is the interruption in the mother-infant relationship by hospitalization and the provision of a stimulating, warm environment with a substitute mother figure. This method achieved success in the four cases reviewed. Berlin, McCullough, Lisha, and Szurek (1957) offer a similar psychoanalytic interpretation in reporting a case study of a 4-yr.-old child hospitalized for 8 mo. at Langley-Porter Clinic. Psychotherapy, involving concomitant counseling to improve the relationship between the parents, led to an alleviation of the child's vomiting reaction.

From the point of view espoused by learning theorists, emesis and rumination may be

¹This study was supported in part by a grant (MH-10993) from the National Institute of Mental Health, United States Public Health Service.

²The authors wish to thank David Kass, the physician in immediate charge of the present case, for giving the authors the opportunity to explore this treatment method and for his assistance during its application. The authors are also indebted to Charles Lobeck, Chairman of the Department of Pediatrics of University Hospitals, Madison, Wisconsin, who made facilities available for use, and to the assigned nursing staff without whose help and cooperation the present result could not have been accomplished. The authors also express their appreciation to Norman Greenfield and Richard Sternbach of the Department of Psychiatry, University of Wisconsin, for the loan of a polygraph, and to Karl G. Stoedefalke of Physical Education for providing additional EMG preamplifiers.

learned habits. In point of fact, vomiting has been clearly demonstrated as a conditioned response in at least three independent studies (Collins & Tatum, 1925; Kleitman & Crisler, 1927; Pavlov, 1927). This prompts the corollary hypothesis that such behavior could be eliminated directly by counterconditioning procedures.

A number of case reports indicate that considerable success may be achieved in modifying alimentary habits in the clinic setting. Both Bachrach, Erwin, and Mohr (1965) and Meyer^a successfully treated adult anorexic patients by making various social and physical reinforcers contingent on eating behavior or weight gain. Lang (1965) described the therapy of a young adult patient who became nauseous and vomited under social stress. In this case, counterconditioning methods increased the patient's tolerance of formerly aversive social situations, and thus markedly reduced the frequency of nausea and emesis.

The only study reviewed, attempting to apply conditioning methods specifically in the treatment of ruminative vomiting was reported by White and Taylor (1967). Electric shock was applied to two mentally retarded patients (23-yr.-old female, 14-yr.-old male) whenever throat, eye, or coughing gestures signaled rumination. They suggest that the shock served to distract the patient and he engaged in other activities rather than ruminating. Significant improvement occurred after 1 wk. of treatment, and gains were maintained at a 1-mo. follow-up.

The following case report illustrates the efficacy of aversive conditioning in reversing the vomiting and rumination of a 9-mo.-old infant whose life was endangered by this behavior. The case is of general interest because of the extreme youth of the patient, the speed of treatment, and the fact that conditioning procedures were undertaken only after other treatments had been either ruled out by diagnostic procedures, or had been given a reasonable trial without success. These data also have further implications for the understanding of aversive conditioning procedures in clinical practice.

^a Meyer, V. Personal communication, 1964.

History of Problem and Family Background

A. T. at the age of 9 mo. was admitted to the University Hospital for failure to retain food and chronic rumination. This infant had undergone three prior hospitalizations for his persistent vomiting after eating and failure to gain weight. Born in an eastern state after an uneventful 39-wk. pregnancy, the patient was bottle fed and gained steadily from a birth weight of 9 lb. 4 oz. to 17 lb. at 6 mo. of age. Vomiting was first noted during the fifth month, and increased in severity to the point where the patient vomited 10–15 min. after each meal. This activity was often associated with vigorous thumbsucking, placing fingers in his mouth, blotchiness of the face, and ruminating behavior. The mother remarked that the start of vomiting may have coincided with her indisposition due to a broken ankle which forced the family to live with maternal grandparents for several weeks. Some friction was reported between the patient's mother and her own adoptive mother concerning care of the child. The patient's father is a part-time college student and the family received financial assistance from the paternal grandfather, a successful dentist. At the time of the most recent hospitalization, the social worker's report suggested that the parents were making a marginal marital adjustment.

Three brief periods of hospitalization which included medical tests (gastrointestinal fluoroscopy, EEG, and neuropsychological testing) failed to find an organic basis for this persistent regurgitation. An exploratory operation was performed and a cyst on the right kidney removed, with no discernible effect on his condition. The patient had no history of head trauma. One previous incident of persistent vomiting in a paternal uncle was noted to be of very short duration. The paternal grandfather and two uncles are reported to suffer ulcers.

Several treatment approaches were applied without success. Dietary changes (Pro-Sobee, skim milk), the administration of antinauseants, and various mechanical maneuvers to improve the feeding situation (different positions, small amounts at each feeding, burping) gave short-lived, if any, relief. As thumb sucking often preceded the response, restraints were tried. However, this did little

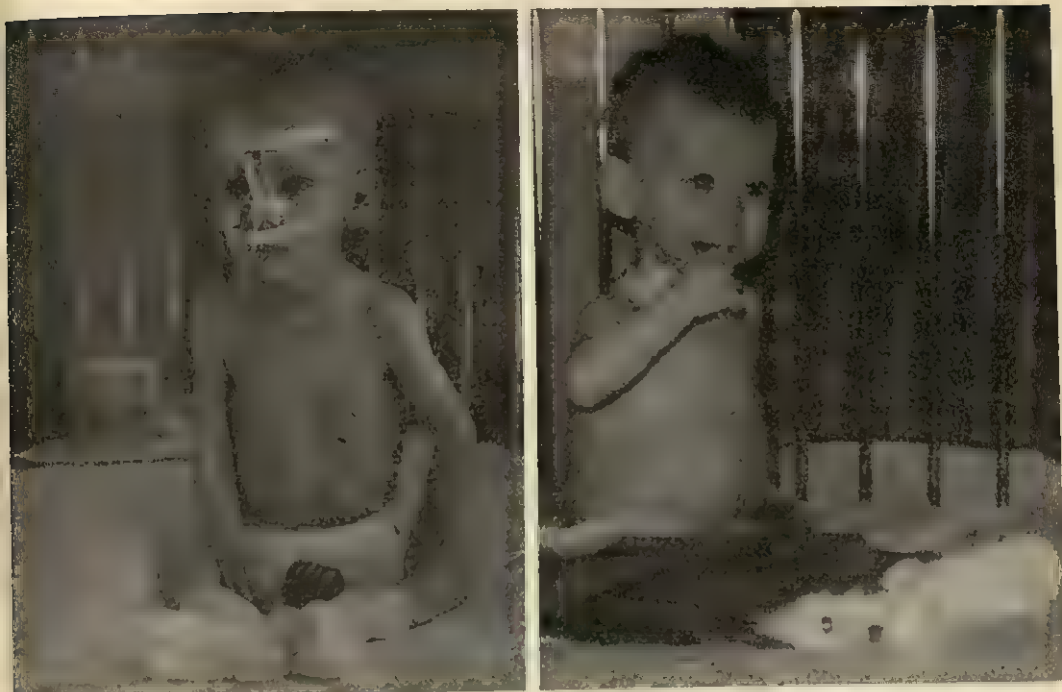


FIG. 1. The photograph at the left was taken during the observation period just prior to treatment. (It clearly illustrates the patient's debilitated condition—the lack of body fat, skin hanging in loose folds. The tape around the face holds tubing for the nasogastric pump. The photograph at the right was taken on the day of discharge from the hospital, 13 days after the first photo. The 26% increase in body weight already attained is easily seen in the full, more infantlike face, the rounded arms, and more substantial trunk.)

to reduce the frequency of emesis. An attempt had been made to initiate intensive nursing care "to establish and maintain a one-to-one relationship and to provide the child with warm, friendly, and secure feelings [nurse's chart]." This had to be abandoned because it was not inhibiting the vomiting and some observers felt that it increased the child's anxiety and restlessness.

At the time the present investigators were called in, the infant was in critical condition, down to a weight of 12 lb., and being fed through a nasogastric pump. The attending physician's clinical notes attest that conditioning procedures were applied as a last attempt, "in view of the fact that therapy until now has been unsuccessful and the life of the child is threatened by continuation of this behavior."

Therapeutic Procedure and Results

The patient was given a private room, continuous nursing care, and assigned a special

graduate nurse to assist in the conditioning procedures. The authors closely observed the infant for 2 days during and after normal feeding periods. He reliably regurgitated most of his food intake within 10 min. of each feeding and continued to bring up small amounts throughout the day. Observers on the hospital staff suggested that vomiting was originally induced by thumb pressure at the back of the throat. However, at this stage thumb manipulations were not a necessary part of the vomiting sequence. He did protest, however, if hand restraint was enforced. His frail appearance and general unresponsiveness, made him a pathetic looking child as seen from a photograph taken just prior to treatment (Figure 1).

In an attempt to obtain a clearer picture of the patterning of his response, electromyograph (EMG) activity at three sites was monitored on a Gilson Polygraph. Responses leading up to and into the vomiting sequence reliably coincided with the nurse's concur-

rent description of the sequence of behavior. Figure 2 illustrates the typical response pattern. The uppermost channel of information represents muscle potentials recorded just under the chin, and shows the sucking behavior which usually preceded vomiting; the lowest channel is an integrated record taken from the throat muscles of the neck; the center channel which monitors the upper chest region is largely EKG artifact. It can be noted from this segment that the onset of vomiting is clearly accompanied by vigorous throat movements indicated by rhythmic, high-frequency, high-amplitude activity, in contrast with quiescent periods and periods where crying predominated.

The authors were concerned with eliminating the inappropriate vomiting, without causing any fundamental disturbance in the feeding behavior of the child. Fortunately, the child did not vomit during feeding, and the sucking and vomiting could be distinguished readily on the EMG. After 2 days of monitoring, conditioning procedures were initiated. The aversive conditioning paradigm called for brief and repeated shock (approximately 1 sec. long with a 1-sec. interpulse interval) as soon as vomiting occurred, con-

tinuing until the response was terminated. An effort was made to initiate shock at the first sign of reverse peristalsis, but not during the preceding sucking behavior. The contingency was determined from the nurse's observations of the patient and the concurrent EMG records. In general, the nurse would signal as soon as she thought an emesis was beginning. If EMG confirmed the judgment, shock was delivered. Occasionally, the EMG would initiate this sequence, with the observational judgment following.⁴ Shock was delivered by means of a Harvard Inductarium to electrodes placed on the calf of the patient's leg. A 3,000-cps tone was temporally coincident with each shock presentation.⁵ Sessions were chosen following feeding to insure some frequency of response. Each session lasted less than 1 hour.

After two sessions shock was rarely required. The infant would react to the shock by crying and cessation of vomiting. By the third session only one or two brief presentations of shock were necessary to cause cessation of any vomiting sequence. Figure 3 illustrates the typical sequence of a conditioning trial.

The course of therapy is indicated in Figure 4. Few shocks were administered after

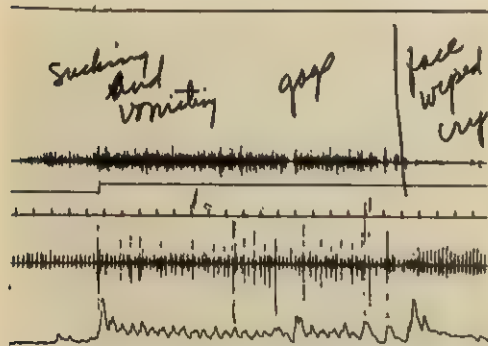


FIG. 2. Three channels of EMG activity are presented. (The nurse observer's comments are written just above the first channel. The intense muscle activity on this line is associated with sucking behavior, recorded from electrodes on the underside of the chin. The second channel is just below the one pulse per second, timing line, and was taken from electrodes on the upper chest, at the base of the throat. The EKG dominates this channel, with some local muscle activity. Electrodes straddling the esophagus yielded the lowest line, which in this integrated record clearly shows the rhythmic pulsing of the vomiting response.)

⁴ Particular thanks are due to Mary Kachoyeanos, the nurse who assisted at all the therapy sessions.

⁵ Shock level was first determined by applying the electrodes to the Es, who judged it to be quite painful and unpleasant. Intensity was incremented slightly during the first and second sessions on the basis of the patient's response, but was subsequently unchanged. The inductarium does not permit for exact or wholly reliable measures of current level. However, under the conditions of treatment described here, the average current was within a range of from .10 to .30 ma., with a cycle frequency of approximately 50 cps. It should be borne in mind that pulses from an inductarium vary widely in amplitude, and the authors' instrument produced some spikes over 10 ma. Electrodes were first applied to the ball of the foot and then moved to the calf for reasons stated in the text. The accompanying tone was generated by a Hewlett-Packard signal generator and administered by a small oval speaker in a free field. The intensity was loud but not painful (approximately 80-95 db.), and varied considerably because of spontaneous changes in the infant's position. It was employed in order to increase the density of the reinforcer and on the possibility that the therapists might employ it alone, if shock proved to have negative side effects.

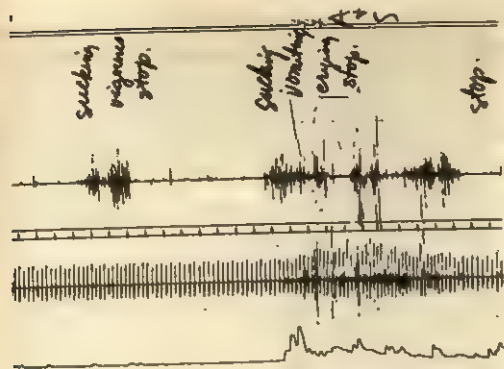


FIG. 3. The electrode positions are the same as in Figure 2. (The top line shows the point at which two brief shocks were administered. It may be noted that they follow closely on the first pulse of the vomiting response and that the rhythmic regurgitation observed in Figure 2 never gets underway.)

the first day of treatment, and both the time spent vomiting and the average length of each vomiting period were abruptly reduced. After only two sessions it seemed that the infant was anticipating the unpleasant consequences of his behavior. He would begin to suck vigorously using his thumb, and then he would remove his thumb and cry loudly.

The data graphed (Figure 4) for the second treatment session represent those reinforcers that the authors are certain were delivered. Early in this session, it became obvious that the infant was not receiving the majority of the administered shocks. The electrodes were at that time attached to the plantar surface of the foot. Observation suggested that the patient had learned to curl his foot, either coincident with emesis or at the first sensation of shock, so as to lift the electrodes off the skin and thus avoid the painful stimulus. At this point, the electrodes were relocated on the calf, and conditioning proceeded normally. If the shock administrations prior to this procedural change are added to those on the graph, Day 3, afternoon figures for emesis period, percentage of emesis, and shock, respectively, are 11 sec., 21.6% and 77.

By the sixth session the infant no longer vomited during the testing procedures. He would usually fall asleep toward the middle of the hour. Figure 5 indicates the sequence of response demonstrating the replacement of

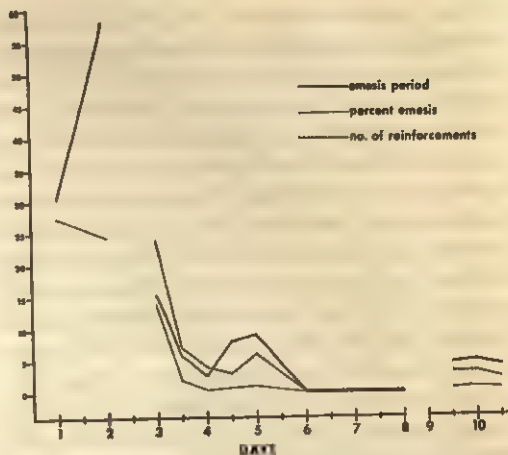


FIG. 4. The abscissa describes successive days (morning and afternoon) on which observation or treatment was accomplished. ("Emesis period" is the length of any continuous period of vomiting. "Percentage of emesis" is the total time spent vomiting divided by the time observed. Sessions varied from 16 min. to 60 min. Treatment began on Day 3 which included two unshocked emesis periods. In Session 10 tone alone was presented on one trial. It is of interest to note that following therapy, nursing staff reported that they could now block the very rare vomiting periods with a sharp handclap.)

vigorous sucking with what the nursing observers described as a "pacifier" use of the thumb.

To vary the conditions under which learning would take place, thereby providing for transfer of effects, the sessions were scheduled at different hours of the day, and while the infant was being held, playing on the

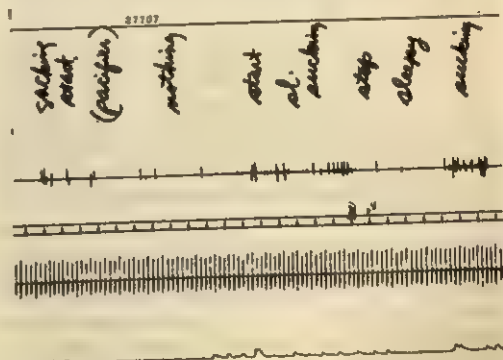


FIG. 5. The above segment is representative of behavior near the end of a conditioning session. (Only mild sucking activity is apparent in the upper EMG channel. The electrode positions are the same as in Figure 2.)

floor, as well as lying in bed. Nursing staff reported a progressive decrease in his ruminating and vomiting behavior during the rest of the day and night, which paralleled the reduction observed across therapy sessions.

After three sessions in which there was no occurrence of vomiting, the procedure was discontinued. Two days later there was some spontaneous recovery, which included some vigorous sucking, with a little vomiting and rumination. Three additional sessions were initiated to maintain the reduced frequency of the response (see Figure 4). Except for a brief slackening prior to these trials, there was a steady, monotonic increase in his weight as shown in Figure 6. In general, his activity level increased, he became more interested in his environment, enjoyed playroom experience, and smiled and reached out to be held by the nurse and other visitors.

The mother was reintroduced the day following the last conditioning trial. She took over some of the patient's caretaking needs, including feedings. There was no marked change in his ruminating behavior at this time. The mother responded well and her

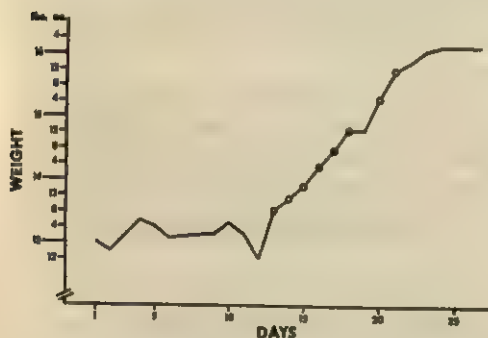


FIG. 6. The infant's body weight as determined from the nursing notes is plotted over time, from well before conditioning therapy was instituted to the day of discharge from the hospital. (Days on which conditioning sessions occurred are marked by circles on the curve. Reinforcers were delivered only on days marked by open circles. The decline in body weight in the few days just prior to therapy was probably occasioned by the discontinuance of the nasogastric pump, in favor of normal feeding procedures. The marked weight gain from Day 13 to 18 is coincident with the first 6 days of therapy. The temporary reduction in weight increase, associated with a resumption of emesis, is apparent at Day 19. The additional conditioning trials appear to have acted immediately to reinstate weight gain.)

child reciprocated her attention. He was discharged from the hospital 5 days later, after exhibiting almost no ruminating behavior. The remarkable contrast in his physical appearance is noted in a photograph taken on the day of discharge (Figure 1).

Follow-up

Correspondence with the mother indicated that there was no further need for treatment. A. T. was eating well and gaining weight regularly. She reported that any thumbsucking or rumination was easily arrested by providing him with other forms of stimulation. He was beginning to seek attention from other people and enjoyed the company of other children. One month following discharge from the hospital, he was seen for a physical check-up. He appeared as a healthy looking 21-lb. child and, aside from a slight anemic condition, was found fully recovered by the attending physician. His local physician reported on a visit 5 mo. later when his weight was 26 lb., 1 oz. "His examinations were negative for any problems. . . . He was eating quite well . . . no vomiting had recurred. He was alert, active and attentive." A snapshot taken by the mother a few weeks before this examination is reproduced in Figure 7. One year after treatment he continues to thrive. Mother and father are both pleased with his development, and no further treatment is indicated.

Discussion

The rapid recovery of this 9-mo.-old male infant following brief aversive conditioning therapy, argues for the effectiveness of behavioral modification in the treatment of this type of psychosomatic disorder. The vomiting and ruminating were treated as maladaptive behavior patterns, and electric shock was used to inhibit a previously well-established response sequence. Elimination of the response was accompanied by increase in the infant's responsiveness to people, as well as substantial weight gains, and physiological improvement.

Treatment was undertaken without analysis of the disorder's antecedents. Nevertheless, the family history of the infant could



FIG. 7. The patient 5 mo. after treatment.

be construed as consistent with other cases in the literature. One clinical worker suggested that a feeling of hostility dominated this infant's home. It is true that the parents' wedding was attended by difficulties and the subsequent birth of the patient occurred before the parents were fully prepared for this responsibility. Furthermore, the mother later expressed anxiety about her marriage and complained of the problem of balancing the separate demands of father and child. She also reported her feeling that her own step-mother had not provided a good maternal model. As a consequence she felt inadequate herself and uncertain in the role.

The caseworker's notes are thus rich in "dynamics," and while one is unable to establish the relative accuracy or significance of

these statements, it is clear that this case is interpretable within traditional personality theories. Nevertheless, therapies generated by this orientation were not successful in the present case. In deference, it should also be noted that "one-to-one" care was not maintained as long or as consistently as in many cases reported in the literature, and despite evidence of some marital discord, no extensive counseling of the parents was undertaken. However, like many psychiatric treatments, the above are expensive of professional personnel and prolonged in duration. The aversive conditioning procedures used here achieved success in little more than a week, and considering the developing danger to the child's life, speed was of more than usual importance.

No evidence of "symptom substitution" was observed following treatment. On the other hand, positive social behavior increased coincident with the successful conditioning therapy. The infant became more responsive to adults, smiled more frequently, and seemed to be more interested in toys and games than he had been previously. An analogous improvement in social behavior was noticed in the defective adults treated by White and Taylor (1967). Lovass, Freitag, Gold, and Kassorla (1965) and Lovaas, Schaeffer, and Simmons (1965) have cited similar effects following the avoidance conditioning of tantrum behavior in autistic children. The latter investigators suggest that the *E*s attained secondary reinforcing value because of their association with shock reduction. This provides the basis for training the children to exhibit affectionate patterns toward adults. In the present case this contingency was very imprecise, and it is not clear that the above mechanism mediated change. What could be called normal infant behavior increased regularly, as the emesis decreased. The social environment appeared simply to replace ruminating as the infant's focus of attention.

Aversive conditioning has been applied widely in adult therapy as well as with autistic children. Eysenck and Rachman (1965) and Feldman (1966) describe its use in treating alcoholic and sexual disorders. However, one hesitates to interpret these findings in a

straightforward manner. Adult patients may submit to aversive conditioning procedures from a variety of motives, and cognitive factors may blunt the impact or distort the meaning of aversive stimuli. The present case is of particular interest because these procedures were successful in treating an apparently normal child. Furthermore, the absence of language and the limited cognitive development achieved at this age permit one to interpret this change as avoidance conditioning, unmitigated by the above factors.

Finally, it should be noted that the present case represents a productive use of psychophysiological recording in therapy. Not only did the EMG provide extensive documentation of the response, but concurrent recording was of considerable help in guiding the treatment effort. Specifically, these records confirmed in an objective manner external observations of mouth and throat movements which seemed to precede emesis. Furthermore, they extended these observations, helping the authors to specify those aspects of the response which were unique to the vomiting sequence, thus assuring that shock was never delivered following noncontingent behavior. Finally, observation of the recordings during therapy probably reduced the latency of reinforcement, particularly during the early trials when the validity of external signs seemed less certain, and provided the clearest indicator of the end of the response when shock was promptly terminated. While the importance of this information to the results obtained cannot be unequivocally established, it certainly increased the confidence of the therapists in their method, and, in turn, the speed and precision with which they proceeded. The further exploration of physiological analysis in the therapeutic setting is encouraged.

REFERENCES

- BACHRACH, A. J., ERWIN, W. J., & MOHR, J. P. The control of eating behavior in an anorexic by operant conditioning techniques. In L. P. Ullmann & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- BERLIN, I. N., McCULLOUGH, G., LISHA, E. S., & SZUREK, S. Intractable episodic vomiting in a three-year old child. *Psychiatric Quarterly*, 1957, 31, 228-249.
- BORISON, H. L. Effect of ablation of medullary emetic chemoreceptor trigger zone on vomiting response to cerebral intra-ventricular injection of adrenaline, apomorphine and pilocarpine in the cat. *Journal of Physiology*, 1959, 147, 172-177.
- COLLINS, K. H., & TATUM, A. L. A conditioned salivary reflex established by chronic morphine poisoning. *American Journal of Physiology*, 1925, 74, 14-15.
- EYSENCK, H. J., & RACHMAN, S. *The causes and cures of neurosis*. San Diego, Calif.: Knapp, 1965.
- FELDMAN, M. P. Aversion therapy for sexual deviations: A critical review. *Psychological Bulletin*, 1966, 65, 65-79.
- KLEITMAN, N., & CRISLER, G. A quantitative study of the conditioned salivary reflex. *American Journal of Physiology*, 1927, 79, 571-614.
- LANG, P. J. Behavior therapy with a case of nervous anorexia. In L. P. Ullmann & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- LOVAAS, O. I., FREITAG, G., GOLD, V., & KASSORLA, I. Experimental studies in childhood schizophrenia: Analysis of self-destructive behavior. *Journal of Experimental Child Psychology*, 1965, 2, 67-84.
- LOVAAS, O. I., SCHAEFFER, B., & SIMMONS, J. Building social behavior in autistic children by use of electric shock. *Journal of Experimental Research in Personality*, 1965, 1, 99-109.
- PAVLOV, I. P. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Lecture III, Oxford, England; Oxford University Press, 1927.
- RICHMOND, J. B., EDDY, E., & GREEN, M. Rumination: A psychosomatic syndrome of infancy. *Pediatrics*, 1958, 22, 49-55.
- WHITE, J. D., & TAYLOR, D. Noxious conditioning as a treatment for rumination. *Mental Retardation*, 1967, 5, 30-33.

(Received September 11, 1967)

INFLUENCE OF THE PARENTS AND VERBAL REINFORCEMENT ON THE PERFORMANCE OF SCHIZOPHRENIC PATIENTS¹

AMERIGO FARINA, JULES D. HOLZBERG,² AND ROBERT R. DIES³

University of Connecticut and Connecticut Valley Hospital, Middletown

Two studies concerned with the psychological deficits displayed by schizophrenic patients were conducted. The first study evaluated the role of the parents and the second measured the effects of praise and censure. For each study, Ss were two groups of male schizophrenic patients, one rated as Good and the other as Poor by means of the Phillips Scale, and a group of nonschizophrenic psychiatric patients. Large differences were found between patients whose parents displayed high and those whose parents displayed low degrees of conflict toward them just prior to the task performance. Verbal rewards and punishments did not appear to affect the patient's behavior and for neither study were the three types of Ss differentially influenced by the manipulations. The main conclusions are that the phenomenon of psychological deficits is extremely complex and that schizophrenic patients are less readily influenced by affective experiences than prior researchers have suggested.

A very large number of studies demonstrate that schizophrenic patients display psychological deficits (Buss & Lang, 1965; Lang & Buss, 1965), defined as inadequacies in the behavior of such patients in comparison to control Ss or in relation to expectations based upon education and intelligence. Usually measured in terms of performance on laboratory tasks, these deficits are reported with such regularity that they would appear to be one of the more demonstrable and stable characteristics of people labeled as schizophrenic. The researchers who have done this work have been concerned with a variety of questions such as whether schizophrenic patients are deficient because they are not responsive to external stimulation (Slechts, Gwynn, & Peoples, 1963) or because they have limited ability to think abstractly (McGaughran & Moran, 1956). The two studies reported in the present paper are related to two of the many existing questions about this phenomenon: (a) what role do other people, and especially the patients' parents, play in these deficits, and (b) in what way are the deficits affected by censure and praise? Relevant research will be reviewed briefly.

¹ This study was financed by a research grant from the National Institute of Mental Health, United States Public Health Service (M-6167).

² Now at Wesleyan University.

³ Now at the University of Kentucky.

When the tasks assigned have been varied with respect to presence or absence of human contents, the presence of such contents has been found to significantly increase the deficits displayed by schizophrenic patients (Brotsky, 1961; Davis & Harrington, 1957; Whiteman, 1954). As these findings imply, the physical presence of other persons also seems to cause schizophrenic patients to perform tasks less adequately than when people are absent (Felice, 1961; Fischer, 1963; Schooler & Spohn, 1960). What is especially pertinent for this investigation, though, is that a number of studies have indicated that the parents play a particularly important role in the elicitation of behavioral deficits. Thus, Dunham (1959), Dunn (1954), Kreinik (1959), and Turbner (1961) have all demonstrated that some of the deficits appear to be *specifically* associated with cues of parents or parents in certain kinds of activities. The most direct study of this association between parental presence and defective behavior, however, is that carried out by Farina and Dunham (1963). These authors induced a group of schizophrenic patients to engage in a rather conflictful interaction with their parents and immediately after this had them perform a task in the presence of cues of the parents. When their performance was compared to that of a group of schizophrenic

patients who had not seen their parents for at least a week, it was found to be significantly poorer. One of the two studies reported here is quite similar to the Farina and Dunham (1963) study but employs a number of methodological improvements including a control group of nonschizophrenic patients.

The second study is closely related to the first. Other groups of schizophrenic and control Ss were administered a preliminary task, after which their performance was either praised, censured, or not evaluated. They were subsequently assigned the same tasks as were given Ss of the first study. Thus, this investigation was designed to facilitate the interpretation of the results of the first study. If the expected deficits were manifested by patients following interaction with the parents, the second study might provide meaningful cues about the specific facets of the interaction which led to such deficits. For example, assuming the results of the second study indicated that censure led to inadequate performance, the findings would argue that it was parental disapproval and criticism which were responsible for the patient's poor performance in the first study.

The second investigation was also designed to provide data which might potentially bring some clarity to the confusing array of results regarding the effect of praise and censure on performance of schizophrenic patients. This literature, which has been reviewed recently (Buss & Lang, 1965), contains reports of censure leading to decrements (Garmezy, 1952; Webb, 1955) as well as to improvement in adequacy of task performance (Atkinson & Robinson, 1961; Cavanaugh, Cohen, & Lang, 1960). However, the nature of the tasks used in these contradictory studies has varied considerably. Thus, Garmezy (1952) had patients discriminate tones varying in pitch while Atkinson and Robinson (1961) required their Ss to learn lists of paired adjectives. Since it has been shown that for normal Ss the effectiveness of verbally administered reward and punishment changes as a function of the complexity of the task (Meyer & Offenbach, 1962), it may be that some of the discrepant results are due to differences in tasks used. It is partly for this reason that in the present studies two tasks

were used, one being an extremely simple reaction-time (RT) task and the other a complex anagrams task requiring the construction of new words out of a given word.

There is at least one more variable which might conceivably account for some of the inconsistent results. Since the category of patients diagnosed as schizophrenic is recognized to be an extremely heterogeneous one, Ss of previous studies may have differed with regard to important personal characteristics. This variability among patients has led to the identification of a number of dimensions, such as process-reactive and good-poor premorbidity and to the discovery that patients differing with respect to these dimensions behave in distinctly different ways (Herron, 1962). For this reason, the two studies used two groups of schizophrenic patients, one group having had a good premorbid history (Goods) and the second group a history of inadequate interpersonal relationships (Poors) as determined by a Phillips Scale rating (Phillips, 1953). This dimension was used since it is objectively assessed and it is highly correlated with other dimensions for categorizing schizophrenic patients (Garfield & Sundland, 1966). The finding that Goods and Poors respond differently to praise and censure would imply that some of the existing confusion regarding psychological deficits may be due to differences in patient characteristics in these contradictory studies.

STUDY 1

Method

The Ss of this investigation were three groups of hospitalized male psychiatric patients and their biological parents. These Ss are described in detail in earlier reports (Farina & Holzberg, 1967; Farina & Holzberg, 1968). All Ss had to meet a series of rather stringent criteria, and there were no significant group differences for age, education, family size, or length of time the parents had been married. One group ($n = 24$) consisted of nonschizophrenic psychiatric patients, while all patients in the remaining two groups were diagnosed as schizophrenic, one of these groups ($n = 26$) being composed of patients rated as Goods while patients in the last group ($n = 24$) were classed as Poors.

For each group, half the patients performed the tasks before interacting with their parents (TBI condition) while the remaining half performed these immediately after the interaction (TAI condition). The interaction procedure is fully described in the

earlier reports. In brief, the mother, son, and father were individually asked how they believed six hypothetical problem situations involving parents and children should be resolved. The three members were then brought together and asked to indicate how they would resolve the same problems as a family. Both the individual and joint sessions were tape-recorded, and the recordings were subsequently used to obtain measures of conflictual behavior.

The first task given the Ss, the anagrams task, required them to make as many new words as possible out of a given word within a 2-min. time period. They were given an initial practice word, "table," and, if necessary, were helped to make new words such as "late" and "ale." Following the practice word, they were presented with the words "chair," "bossy," "scolding," "mother," and "father," in the order listed. A record was made of all new words as well as the time in the 2-min. time period when each was produced. The second task measured simple RT and required visual motor skills in contrast to the thinking necessary to perform the anagrams task. After a standard practice period, each patient was given seven trials at each preparatory interval of 1, 2, 4, 8, and 16 sec., all seven trials at the 1-sec. interval being given first and then followed by the other intervals in the order listed.

Results

The results obtained are summarized in Table 1. The patients' productions were classed as words if the word appeared in Funk & Wagnall's Standard College Dictionary, 1963 Edition, otherwise they were scored as errors. As may be seen, the differences among groups and between conditions are quite small, not particularly consistent, and seemingly uninterpretable. Other measures not presented in Table 1, such as the number of words produced for each specific word, number of three-letter words, etc., show a pattern quite similar to that of Table 1. These results were analyzed by means of three-factor (groups, condition, and word or preparatory interval) analyses of variance as described by Winer (1962, pp. 337-349). There were no groups of condition main effects nor were any of the condition interactions with the other variables significant. Only one significant interaction involving the groups factor was found, and it was a Groups \times Preparatory Interval interaction for RT. The Poors were relatively slower than the Goods and Controls on the 2- and 4-sec. preparatory intervals. Additional analyses revealed that in doing the anagrams task the Controls ar-

TABLE 1
PERFORMANCE ON ANAGRAMS AND REACTION-TIME TASKS BY EACH GROUP OF SUBJECTS AS A FUNCTION OF CONDITION (STUDY 1)

Group	Anagrams task		Reaction-time task
	Mean no. words produced to each word presented	Total no. errors made to all words presented	Mean time per trial in 1/100ths of a second
Controls			
TBI ^a	9.2	8.2	20.2
TAI ^b	8.3	7.3	18.9
Goods			
TBI	7.1	3.5	20.0
TAI	7.5	6.3	21.1
Poors			
TBI	8.4	5.2	22.7
TAI	7.6	5.3	21.1

^a Performed task *before* interacting with parents.

^b Performed task *after* interacting with parents.

rived at the first word significantly faster than the Goods and Poors and that they produced the last word significantly later in the 2-min. period than the Goods and somewhat later than the Poors. Thus, the Control group Ss appear to have worked longer than the schizophrenic patients. No other differences were found.

Although these results suggest that the interaction with the parents did not influence the performance on the two tasks, it is still possible that some of the patients were affected by the experience. Conceivably the interaction was benign for some of the patients but traumatic for others because of the difference in degree of conflict they encountered in interacting with the parents. It would seem reasonable to expect that traumatized patients would manifest more deficits in their performance than other Ss. To test this hypothesis, 37 Ss in the TAI condition were ranked in terms of the amount of conflict their parents had expressed toward them during the interaction.⁴ Eliminating the pa-

⁴ The indexes used were: *failure to agree* (when the mother and/or father disagreed with the son), *interruptions* (when the mother or father interrupted the son), and *disagreements and aggressions* (displayed by either parent toward the son). Further details about the indexes are available in the earlier publications.

tient who was at the median, high, and low parental conflict groups of 18 Ss each were formed. Analyses of variance revealed that the high-conflict group made significantly more errors in the anagrams task ($F = 10.1$, $df = 1/170$, $p < .01$) and produced significantly more words ($F = 21.4$, $df = 1/170$, $p < .01$) than the low-conflict groups. Since these two dependent measures, errors and words, are only slightly correlated ($r = .105$), these findings suggest that a conflictual interaction with the parents causes patients to make more errors or to produce more words without changing the proportion of errors. There was no relationship between group membership and either tendency to produce more words or to make more errors. No differences of any sort were found for the reaction time results.

STUDY 2

Method

The Ss of the second study were also three groups of hospitalized male psychiatric patients, and, as in the former study, one group was composed of non-schizophrenic patients, the second of Good, and the third of Poor premorbid schizophrenic Ss. There were 30 patients in each group, and parents were not involved in this study. These Ss had to satisfy the same criteria as those in the first study except the one requiring that the parents be alive, in good health, and living together. These criteria are described in the earlier publication (Farina & Holzberg, 1967). Groups were comparable for education, occupation, and number of hospital admissions. The Goods were significantly older than the Controls and Poores, however, while the latter two groups did not differ from each other. In view of the results, this age difference does not present a problem. They appear to be reasonably representative of relatively acute state hospital patients, being about 30 yr. of age on the average and of low socioeconomic status.

All Ss meeting the criteria were taken to an experimental room and given two tests. One was the WAIS Vocabulary subtest and the other was a measure of how rapidly the S could tap a standard telegraph key. The latter task required S to tap as rapidly as possible for 30 sec. after a 5-sec. practice period, the number of taps being automatically counted by a Veeder Roots counter. These tasks served as a pretext for the experimental manipulation which consisted of praising or censuring the Ss. A control group received no evaluation. One-third of each group of Ss, Goods, Poores, and nonschizophrenic, were randomly assigned to each of the experimental conditions.

Following the two initial tasks, patients in the

praise condition were told that they appeared to have done very well but that the E would score the records and then inform them. The patients were allowed to wait for 3 min. while the protocols were apparently scored and they were then told they had done much better than anticipated and that they were indeed very good at those tasks. They were then instructed they would do some other things which were similar, and the RT and anagrams tasks were administered. These tasks were identical to those of the first study except that three additional words, "picture," "kitchen," and "curtain," were included in the anagrams task. These were given after the word "chair." In the censure condition the same procedure was followed but the patients were told they appeared not to have done very well and, after the 3-min. wait, that they had done very poorly and were very poor at these tasks. The nonevaluated group Ss were made to wait 3 min. after being told that the E needed to do something unrelated to the study, and no information about their performance was given. This is substantially the same procedure which was used by Webb (1955), who reported large differences in task performance between his groups of praised and censured schizophrenic patients. For each group, 5 of the 10 Ss in each condition received the RT task first, and the remainder received the anagrams task first.

Results

Results were examined by means of four-factor analyses of variance in which three of the variables were between Ss and one was within Ss. The three between-Ss variables were groups (Goods, Poores, and Controls), evaluation (praise, censure, and nonevaluation), and order of tasks (RT first and anagrams first). The within-Ss variable was either preparatory interval or word depending on which of the two measures was being examined. Results of these analyses were consistent and unambiguous. For number of words produced in the anagrams task and for time required to respond in the RT task there were significant main effects for groups. The F scores were 3.36 and 3.56, respectively, which for $df = 2$ are statistically significant ($p < .05$). As may be seen from the mean scores presented in Table 2, the Controls produce more words and react faster than the Ss of the other groups, while Poores produce the least number of words and are the slowest. No other differences were found to be significant for either measure except for main effects for preparatory interval for RT and words for the anagrams task which are to be ex-

TABLE 2

PERFORMANCE ON ANAGRAMS AND REACTION-TIME TASKS BY EACH GROUP OF SUBJECTS AS A FUNCTION OF TYPE OF EVALUATION (STUDY 2)

Group	Anagrams task: Mean no. words produced to each word presented	Reaction-time task: Mean time per trial in 1/100ths of a second
Controls		
Praised	9.10	19.07
Censured	9.28	18.70
Nonevaluated	8.88	19.49
Total group average	9.08	19.09
Goods		
Praised	7.89	23.79
Censured	6.95	21.26
Nonevaluated	8.00	20.54
Total group average	7.61	21.86
Poors		
Praised	7.88	20.82
Censured	7.34	25.31
Nonevaluated	6.75	27.49
Total group average	7.32	24.54

pected and are not particularly informative. Two other dependent measures from the anagrams task, number of errors and time required to produce the first word, were similarly analyzed. For these scores, only the main effect for words was found to be significant. Thus, the results indicate that neither praise nor censure had any effect on the three groups of Ss nor did the groups perform differently as a function of the task administered.

DISCUSSION

The clearest implication of the present findings is that the phenomenon of behavioral deficits in schizophrenia is an extremely complex one. It was found that the performance of high- and low-conflict Ss differed on the anagrams but not on the RT task, and that patients exposed to high parental conflict made more errors and produced more words than low-conflict Ss. Moreover, while this marked effect was found as a function of the nature of the interaction, the seemingly more important variable was that of being exposed or not being exposed to a conflictual interaction with the parents prior to

the task. This did not appear to meaningfully influence the performance. These findings, then, do indicate that such factors as the nature of the tasks used, the kinds of dependent measures used, and the nature of the manipulations to which patients are exposed significantly affect results. For example, on the basis of the present study one might conclude that exposure to the parental interaction had either a facilitating or a debilitating influence upon the behavior of schizophrenic patients depending on the dependent measures considered. A recent study by Lupino (1966) further illustrates the complexity of this research area. Luppino showed that praise, administered the same number of times in the same way by the same E for doing the same task, can lead to very different performance depending on whether the easy or difficult parts of the performance are praised. The general conclusion to which this leads is that it is very premature at the present time to attempt to infer the nature of schizophrenia from studies of psychological deficits as is frequently done (Buss & Lang, 1965; Lang & Buss, 1965). Additional and much more sophisticated studies should precede such an effort if we are not to be badly misled.

Some of the findings of the present study are relevant to much contemporary theorizing about the nature of schizophrenia and its etiology. Such theorizing supposes that the family, and especially the mother, plays an important role in this condition and that the parent-child relationships have been more disturbing for schizophrenics than for non-schizophrenics (e.g., Bateson, Jackson, Haley, & Weakland, 1956; Coleman, 1956, pp. 276-285). In contrast to these assumptions, the first experiment demonstrated that schizophrenic patients encountered their parents under quite realistic and potentially traumatizing conditions without showing any obvious disturbances. The degree of conflict experienced in the parental interaction greatly influenced the patients' behavior, but high conflict did not lead to unambiguous deterioration of behavior as this kind of theorizing would predict. It was also found that there were no differences between schizophrenic and control Ss in their response to parental inter-

action (Table 1) nor did these groups react differently to the manipulation used in the second study. These negative findings may be due to the fact that in the present study the control groups consisted of nonschizophrenic psychiatric patients whereas in other studies the control Ss have typically not been psychiatric patients. While a more typical control group might have yielded results more similar to previous ones, the present findings raise the possibility that the deficits which have been assumed to be specific to schizophrenic patients may also be characteristic of other psychiatric populations.

Although the only variable found to be associated with changes in psychological deficits was degree of parental conflict expressed, group differences were found which are quite consistent with earlier studies. In the second study, the two groups of schizophrenic Ss were slower to react in the RT task and produced fewer words in the anagrams task than the Controls, and the Poors performed less adequately than the Goods. While there are some discrepancies between the findings of the first and second studies, even in the first study the only significant group difference obtained indicated the Poors to be relatively less adequate at some preparatory intervals in comparison with Ss of the other two groups. A number of other studies which, like the present ones, indicate the nonschizophrenic Ss to be most adequate, the Goods next, and the Poors least adequate in the performance of various tasks have been reported (Buss & Lang, 1965; Garmezy & Rodnick, 1956; Lang & Buss, 1965). Thus, the present results, in agreement with the preceding ones, indicate that behavioral deficits are an important facet of psychopathology and merit further research attention. It should also be noted that these group differences in performance imply that the tasks used are valid measures of psychological deficits. Although it is possible that other procedures might yield much larger group differences than the ones used in these studies, the results obtained prompt the conclusion that schizophrenic patients are considerably less influenced by affective experience than prior researchers have suggested.

REFERENCES

- ATKINSON, R. L., & ROBINSON, N. M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, 62, 322-326.
- BATESON, G., JACKSON, D. D., HALEY, J., & WEAKLAND, J. H. Toward a theory of schizophrenia. *Behavioral Science*, 1956, 1, 251-264.
- BRODSKY, M. J. Interpersonal stimuli and interference in a sorting task. *Dissertation Abstracts*, 1961, 22, 2068.
- BUSS, A. H., & LANG, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *Journal of Abnormal Psychology*, 1965, 70, 2-24.
- CAVANAUGH, D., COHEN, W., & LANG, P. J. The effect of "social censure" and "social approval" on the psychomotor performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1960, 61, 213-218.
- COLEMAN, J. C. *Abnormal psychology and modern life*. Chicago: Scott, Foresman, 1956.
- DAVIS, R. H., & HARRINGTON, R. W. The effect of stimulus class on the problem-solving behavior of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1957, 54, 126-128.
- DUNHAM, R. M. Sensitivity of schizophrenics to parental censure. Unpublished doctoral dissertation, Duke University, 1959.
- DUNN, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *Journal of Personality*, 1954, 23, 48-64.
- FARINA, A., & DUNHAM, R. M. Measurement of family relationships and their effects. *Archives of General Psychiatry*, 1963, 9, 64-73.
- FARINA, A., & HOLZBERG, J. D. Attitudes and behaviors of fathers and mothers of male schizophrenic patients. *Journal of Abnormal Psychology*, 1967, 72, 381-387.
- FARINA, A., & HOLZBERG, J. D. Interaction patterns of parents and hospitalized sons diagnosed as schizophrenic. *Journal of Abnormal Psychology*, 1968, 73, 114-118.
- FELICE, A. Some effects of subject-examiner interaction on the task performance of schizophrenics. Unpublished doctoral dissertation, Temple University, 1961.
- FISCHER, E. H. Task performance of chronic schizophrenics as a function of verbal evaluation and social proximity. *Journal of Clinical Psychology*, 1963, 19, 176-178.
- GARFIELD, S. L., & SUNDLAND, D. M. Prognostic scales in schizophrenia. *Journal of Consulting Psychology*, 1966, 30, 18-24.
- GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality*, 1952, 21, 253-276.
- GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implica-

- tions for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease*, 1959, 129, 450-466.
- HERRON, W. G. The process-reactive classification of schizophrenia. *Psychological Bulletin*, 1962, 59, 329-341.
- KREINIK, P. S. Parent-child themes and concept attainment in schizophrenia. Unpublished doctoral dissertation, Duke University, 1959.
- LANG, P. J., & BUSS, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology*, 1965, 70, 77-106.
- LUPPINO, A. V. The nature of chronic schizophrenia: Implications derived from the effects of social reinforcements on performance. Unpublished doctoral dissertation, University of Connecticut, 1966.
- MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object sorting behavior of schizophrenic and nonpsychiatric groups. *Journal of Abnormal and Social Psychology*, 1956, 52, 43-50.
- MEYER, W. J., & OFFENBACH, S. I. Effectiveness of reward and punishment as a function of task complexity. *Journal of Comparative and Physiological Psychology*, 1962, 55, 532-534.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- SCHOOLEER, C., & SPOHN, H. E. The susceptibility of chronic schizophrenics to social influence in the formation of perceptual judgments. *Journal of Abnormal and Social Psychology*, 1960, 61, 348-356.
- SLECHTA, J., GWYNN, W., & PROPLES, C. Verbal conditioning of schizophrenics and normals in a situation resembling psychotherapy. *Journal of Consulting Psychology*, 1963, 27, 223-227.
- TURBINER, M. Choice discrimination in schizophrenic and normal subjects for positive, negative, and neutral affective stimuli. *Journal of Consulting Psychology*, 1961, 25, 92.
- WEBB, W. W. Conceptual ability of schizophrenics as a function of threat of failure. *Journal of Abnormal and Social Psychology*, 1955, 50, 221-224.
- WHITEMAN, M. The performance of schizophrenics on social concepts. *Journal of Abnormal and Social Psychology*, 1954, 49, 266-271.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received June 19, 1967)

ing minimal trance depth despite concerted efforts to induce hypnosis. The *Ss*, at best, were able to respond only to ideomotor suggestions such as arm levitation and eye closure, attaining a score of $\leq +2$ on the rating scale of trance depth. The mean and standard deviation of their scores, respectively, on the SHSS:C, were 1.79 and 1.48. Two *Ss* had experience in simulating from other studies.

Procedure

In a prior independent experiment, *Ss* met *E*₁ who was to serve as the nonblind *E* in the present study. He scheduled *Ss* to take part in the present study, and administered initial instructions to hypnotic and simulating *Ss*.³ All *Ss* were told that the study with the author (*E*₂) would investigate the effect of hypnosis on a variety of tests, and that the tests would be given on 2 successive days. Simulating *Ss* were instructed, additionally, that they were to simulate the part of a hypnotized *S* and behave just as they thought a deeply hypnotized *S* would behave. They were asked, explicitly, not to disclose the fact that they were unaffected by hypnosis and to conceal the fact that they had not been hypnotized until *E*₁ saw them after the experiment had been completed. To motivate simulators for their task of deception *E*₁ told *Ss* that *E*₂ would stop the experiment if he caught on they were faking. If *E*₂ continued the experiment *Ss* were to assume that they were handling the deception well. Even if they thought they had made an error they were told to continue. It was stressed that intelligent *Ss* are able to handle the deception successfully.

The study was divided into two sessions and all *Ss* were run in each of three conditions. Condition I established a base control set of personality measures. Condition II investigated the effect of the artificially induced conflict on personality functioning when *Ss* were amnesic for the source of the conflict, and Condition III examined the effect on personality functioning of lifting amnesia for the complex.

First Session

Condition I. Trance was induced in a hypnotic session which was oriented personally to individual *Ss*; *E* varied the induction procedure and hypnotic testing to accord with the responsiveness of the *S*. Although the form of testing differed slightly among *Ss*, the following items were administered in a standard order (as given): arm levitation, eye closure, arm rigidity with challenge, experience of cold and nose itch, fantasy experience with suggested dream

episode, delusion about a negatively hallucinated number (tested by the inability to handle arithmetical computations), age regression with positive visual hallucination, posthypnotic suggestion, and suggested amnesia for all trance events.

After awakening, *E* tested amnesia and the posthypnotic suggestion and reinduced trance if either suggestion was not met successfully.⁴ After success on these items a battery of personality tests was administered to *Ss* in the waking state to complete the session. The battery contained Card 3 BM of the Thematic Apperception Test (Murray, 1943), the Rotter Incomplete Sentences Blank, College Form (Rotter & Raskin, 1950), the Semantic Differential Test (Osgood, Suci, & Tannenbaum, 1957), and a word association test which was devised by the *E*. The Semantic Differential and word association tests were designed especially to relate to the complex.

The Semantic Differential Test employed 10 scales and 10 concepts. Concepts were: "child," "honesty," "hypnosis," "love," "me," "my ideal self," "my mother," "peace of mind," "school," and "self control." Scales were selected to represent Osgood's three main factors of Evaluation (F1), Activity (F2), and Potency (F3). Scales representing each factor loaded highly on that factor and low on the remaining two factors. F1 was represented by "clean-dirty," "good-bad," and "kind-cruel"; F2 by "active-passive," "excitable-calm," and "hot-cold"; and F3 by "free-constrained," "hard-soft," and "strong-weak." "Intentional-unintentional" was the tenth scale. Scales and concepts were rotated against each other to minimize halo effects, and scales representing the same factor were alternated in polarity direction to prevent the formation of position preferences. The test was arranged in a booklet of 10 pages each presenting a randomly selected order of the 10 concepts paired with 10 different scales. Throughout the test, concepts were paired only once with each scale.

The word association test listed 40 words. The *S* was instructed to say the first word which came to his mind after he read the stimulus word on a set of cards which were handed to him by *E*. Normative data were available for a sample of 500 male and 500 female American university students (Palermo & Jenkins, 1964) on 23 of the stimulus words. When *S* had completed the test, *E* ended the session and remarked that he would see him the next day for the final phase of the experiment.

³ The independent study investigated the effect of motivating instructions on the hand strength of low and high hypnotizable *Ss*. At the end of the experiment *Ss* were informed what the next study would be. The time between the two studies varied for *Ss*. Two *Ss* went onto the present study immediately while others came back as much as a month later. The average time between studies was 13 days.

⁴ The posthypnotic suggestion was an integral part of a study of posthypnotic behavior outside the experimental setting. The *Ss* were given the suggestion to respond to the cue word "experiment" everytime they heard the word in the next 48 hr. The *Ss* were tested for the response, away from the *E*, by the project secretary as they left the building and as they reentered it on the following day. Details of the study are reported elsewhere (Orne, Sheehan, & Evans, 1968: "Occurrence of posthypnotic behavior outside the experimental setting."

Second Session

Condition II. On the second day a deep trance was established and instructions were given to Ss to induce an anxiety conflict with amnesia for the source of the conflict.⁵ Following these instructions S was given amnesia suggestions for the entire session, awakened and tested for amnesia. Trance was reinduced until S reported no memory of the events of the session. The S then was given the test battery to determine the effect of the induced conflict on personality functioning.

Condition III. After S had finished the battery he was put back into trance and given the following instructions:

When I wake you up now, you will remember the incident I told you about. You will remember the park and finding the red purse. You had forgotten it before but now you remember the incident completely. You recall everything just as it happened to you.

⁵ Conflict instructions to S were as follows: "On your way home from school many years ago, you crossed a park. Remember many years ago as a small child crossing a park on your way home from school. Remember it just as it happened to you; remember now going across a big park, green and round. It was a big park, green and round. Halfway across this big park you look down; you see a bright, red purse. You pick it up and find two shiny, big pennies. Two shiny pennies in a bright, red purse. You carry it with you and on your way home from school you pass a candy store, and you see some candy which you like very much. You are tempted to go in and buy it with the two pennies but you know you shouldn't. Your mother would want you to give the purse to her so that she could find the owner. The temptation, however, is too strong and you go in and buy the candy with the two pennies. When you get home you give the empty purse to your mother and you tell her that there was nothing in it. You feel terrible; you lied to your mother, and you were dishonest about the pennies. You think your mother knew that you were not telling the truth and this makes you feel bad. She would be so hurt to think you are deceiving her. But you keep it all to yourself and don't tell her. As you grow up from childhood, you sometimes feel uneasy about being dishonest and lying to your mother. You've always felt very uneasy about stealing that money from the red purse which you found in the park on your way home from school that day. It is only a small thing but you feel terrible about it and about not having told your mother. You feel very uneasy and anxious but you try to forget it. You're grown up now. You're no longer a little child but grown up now, but you still feel very uneasy about that incident. You have forgotten that time you stole the money and lied to your mother. You have forgotten all about it, but it worries you, makes you feel uneasy."

The S was awakened, asked what he remembered, and given the test battery again. This condition studied the effect of absence of amnesia on personality functioning and is relevant to the important question whether the essential therapeutic step in the relief of neurotic conflict is recovery from amnesia or some other manifestation of dissociation. McDougall (1926) suggests that the revival of emotion, or reliving the emotional experience in itself, does not result in relief. He states that, in cases where relief follows, there is a relief of dissociation, and this is the essential therapeutic step in the alleged abreaction process. What is important in therapeutic cure is "the attempt to reintegrate into consciousness the complex that has become autonomous [p. 462]." In the present study, lifting the amnesia without removing the original source of the conflict allows experimental investigation of this hypothesis. It is suggested that in Condition III Ss will show less evidence of conflict than in Condition II where the complex has been successfully repressed.

At the conclusion of the experiment when the test battery was completed for the third time, S was rehypnotized and told that the story he recalled was only vividly imagined and that the incident never really happened. Amnesia was lifted and S was awakened. A comprehensive interview was then conducted by E₁. The E asked Ss for their comments on the experiment and their subjective reactions to the study. A check was made to establish that hypnotic Ss were in trance throughout the session and that the simulators experienced no trance effects.

RESULTS

Results for 3 Ss in the hypnotic group and 1 S in the simulating group were excluded from analysis.⁶ Fourteen hypnotic Ss and 13 simulating Ss provided data for analysis.

Sentence Completion Test

Two independent raters scored the Rotter Incomplete Sentences Blank for adjustment. Completions were scored by comparing them against typical items in empirically derived scoring manuals for men and women (Rotter & Rafferty, 1950) and by assigning to each response a scale value from 0 to 6. Scores were summed over the 40 items to give a total score which was taken as the test index of adjustment. Tests were coded to preclude S identification, and both scorers were un-

⁶ In the hypnotic group one S was given two sessions without conflict; the order of Conditions II and III was reversed for another S; and a third S failed to develop any amnesia for the source incident. In the simulating group one S refused to simulate and showed evidence of entering light trance.

familiar with the experimental design. The interrater consistency of the raters was examined for each group separately on each condition, and for groups and conditions as a whole. There was essential agreement between test scores for all meaningful comparisons between raters. The minimum correlation between test scores for any comparison was .93; the maximum, .99. Table 1 sets out the means and variances of adjustment scores listed by conditions and groups. Results are shown for the rater who used the wider range of scores. High scores represent greater degrees of maladjustment.

Results in Table 1 were analyzed by Lindquist's Type I split plot model of analysis of variance (Lindquist, 1953). Classifications were for conditions (I, II, III) and *S* grouping (hypnosis, simulating). There was no significant difference in performance between hypnotic and simulating *Ss* over conditions. Condition I characteristically showed better adjustment scores than Conditions II and III. There was a significant main effect for conditions ($F = 7.59$, $df = 2/50$, $p < .01$), but the differences in adjustment scores among conditions were evident for both the simulating *Ss* and hypnotic *Ss*.

Word Association Test

The *Ss*' association responses were analyzed in terms of commonality scores, which sum the proportional frequencies of words as expressed by *Ss* in the normative population, and in terms of new responses which were original to the test sample. An original response was defined objectively as one which

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ADJUSTMENT SCORES FOR CONDITIONS AND SUBJECT GROUPS

Group	Condition		
	I	II	III
Hypnotic			
\bar{X}	136.50	150.07	150.93
<i>SD</i>	23.28	20.44	23.54
Simulating			
\bar{X}	133.69	143.38	145.69
<i>SD</i>	19.92	24.36	24.94

TABLE 2

AVERAGE COMMONALITY SCORES (C) AND ORIGINAL RESPONSES (O) FOR SUBJECT GROUPS AND EXPERIMENTAL CONDITIONS

Condition	Measure	Group	
		Hypnotic	Simulating
I	C	5.47	7.55
	O	3.79	.85
II	C	5.34	6.90
	O	3.31	2.08
III	C	5.47	6.49
	O	2.93	1.92

was not contained anywhere in the set of normative responses. Table 2 lists the average commonality scores and numbers of original responses for the two experimental groups. Results are considered only for the 23 words common to the word lists of Palermo and Jenkins (1964).

Results in Table 2 were analyzed by Lindquist's Type I split plot model of analysis of variance. Classifications were experimental condition (I, II, III) and *S* grouping (hypnosis, simulating). Analysis showed a significant main effect for groups ($F = 4.35$, $df = 1/25$, $p < .05$), and a significant interaction between groups and conditions ($F = 3.67$, $df = 2/50$, $p < .05$). Simulators gave more original responses in Conditions II and III as compared with the control condition, but hypnotic *Ss* gave fewer unusual responses under both conditions. The latter finding is evidence that hypnotic *Ss* performed differently from simulating *Ss*. Under conflict, hypnotic *Ss* inhibited spontaneity of response, while simulators did not. However, in interpreting these differences it is necessary to take into account the characteristics of the faking *Ss* introduced by the act of simulating. Simulators may have felt that, in order to prevent repetition of responses already given in the control condition, it would be appropriate to give original responses in Conditions II and III.

Examination was made of the original responses which were related to the complex. Five raters rated each original response paired with its relevant stimulus. All responses were

coded to prevent identification of *S* grouping and condition. Nineteen responses were selected as complex related which were agreed upon by three of the five raters. Examples of such responses (in italics) and their stimuli are: *RED-dislike*, *BUYING-candy*, *RED-purse*, *AFRAID-dishonest*, *FINGERS-found*, *YOUNGER-scared*, *RED-park*, *BOY-spend*. The average numbers of complex related responses for hypnotic Ss were 1.07 and .79 for Conditions II and III, respectively. Simulators showed an average score of .54 for both conditions. These differences between groups were insignificant, but it is of interest to note that although hypnotic Ss showed less tendency to give unusual responses in Conditions II and III while simulating Ss showed more, they actually gave more responses which were related to the induced conflict.

Semantic Differential Test

Semantic data were analyzed to detect instances of psychopathological perception, and to test for differences in meaning attributed to key concepts by hypnotic and simulating Ss. Data were arranged to represent actual ratings of concepts on scales, and difference scores (*D* scores) between concepts over the total semantic space. The *D* score measured the semantic similarity between concepts, and took the sum of the squared differences between ratings on two concepts for selected conditions or groups, summing over all scores (Osgood et al., 1957).

The Ss' ratings of concepts were averaged over nine scales⁷ and the changes in meaning of specific concepts were examined for hypnotic and simulating Ss. Differences were analyzed between groups for each condition and between conditions for the two groups, separately. Of 90 comparisons made, only three differences were significant (involving the concepts hypnosis, me, and school). Considering the number of test comparisons, the differences obtained can be interpreted reasonably as chance occurrences.

Distances between each concept (*m*) and

every other concept were calculated and then entered in $m \times m$ matrices which represented the semantic structure of the total set of *m* concepts and gave the distance or similarity relations among them. Matrices of scores for real and simulating Ss were correlated for Conditions I, II, and III, separately. Analysis of results showed a high degree of similarity in performance between hypnotic and simulating Ss in each condition. For Conditions I, II, and III the *D*-score matrices of hypnotic Ss and simulators correlated .76, .56, and .58, respectively.

Change scores from one condition to another were correlated for both groups. The change in meaning of concepts from Condition I to II was similar for hypnotic and simulating Ss (for matrices of rating scores, $r = .45$, $df = 88$, $p < .001$), as was also the change in meaning of concepts from Condition II to III (for rating scales, $r = .23$, $df = 88$, $p < .05$). However, the change in meaning from Condition I to III for hypnotic Ss bore no relationship to that shown by the simulating Ss (for rating scores, $r = .09$, $df = 88$, $p > .10$; and for difference scores, $r = .01$, $df = 34$, $p > .10$). The latter finding is evidence that Condition III discriminated hypnotic and simulating Ss.

Thematic Apperception Test (Card 3 BM)

Stories were analyzed to provide comparison of test protocols with normative data on an American population of 50 college students and 100 psychiatric patients (Eron, 1950). Two raters blind to *S* grouping and condition scored Ss on four variables selected by Eron: (a) the emotional tone; (b) emotional outcome of the stories (rated on a 5-point scale ranging from "very happy" to "very sad"); (c) theme, scored according to a comprehensive checklist provided by Eron; and (d) misidentification of the sex of the huddled figure which is ambiguously defined in the test card. Tests were also rated on a 5-point scale measuring complex relatedness.

The emotional tone and outcome of the stories were rated as characteristically unhappy for all Ss. Frequency analysis of themes showed that, for both raters, hypnotic Ss gave different themes in the control condition from

⁷ Summation did not include the scale "intentional-unintentional" since this scale represents a fourth, unknown factor in semantic space and is not independent of the three main factors selected.

TABLE 3

FREQUENCIES OF SAME-SEX IDENTIFICATION OF
AMBIGUOUS FIGURE BY HYPNOTIC AND
SIMULATING SUBJECTS

Group	Condition		
	I	II	III
Hypnotic	2	5	2
Simulating	0	1	1

simulating Ss, but emphasized the same themes in Condition II.^a In Condition II both groups of Ss most frequently reported "guilt-remorse" and "pressure from parents" which are clearly complex related. In Condition III both groups expressed "guilt-remorse," but hypnotic Ss next emphasized "inadequacy" while simulating Ss expressed "exhaustion." The hypnotic group gave three unusual themes in the control condition which were not included in the special list compiled from responses typically given to the card (Eron, 1950), whereas simulators gave no unusual themes. On Condition II, however, hypnotic Ss gave no unusual themes, whereas simulators gave two atypical themes. This reflects the tendency, previously observed, for hypnotic Ss to show reduced spontaneity under induced conflict.

Nearly all Ss immediately identified the ambiguous figure as of the opposite sex. In only 10 instances did Ss identify the figure immediately according to their own sex. Table 3 sets out the frequencies of such occurrences.

Table 3 shows that identification according to S's own sex was characteristic only of the hypnotic group in Condition II. Four of the five Ss who perceived correctly in this condition were observed to respond posthypnotically outside of the experimental setting when tested by the project secretary on a pre-arranged cue (see Footnote 4). The phenomenon of extraexperimental response to a post-hypnotic suggestion is known only to occur with Ss capable of deep trance. The finding suggests that identification with the huddled figure typically perceived as unhappy and

with little chance of escaping the conflict was associated with trance depth. Hypnotic Ss, who were markedly susceptible to trance, identified more with the figure under induced conflict than did other Ss.

An analysis of variance (Lindquist's Type I design) carried out on the ratings of complex relatedness^a showed significant differences among conditions across Ss ($F = 10.40$, $df = 2/50$, $p < .001$), but there was no differential effect for hypnotic Ss. There was some evidence that Condition III discriminated Ss. Eighty-six percent of the hypnotic Ss and 85% of the simulating group were rated as having given complex related responses in Condition II; in Condition III, 71% of the hypnotic Ss continued to give related responses, but only 46% of the simulators showed any influence of the complex in their stories. Some simulating Ss apparently expected that the complex should be removed in Condition III, but hypnotic Ss characteristically behaved uniformly under both conditions of conflict.

At the end of Session I on the first day, E_2 rated Ss on whether he thought they were simulators or true hypnotic Ss. The Ss' responses were analyzed in terms of E 's perceptions to detect the possible effect of E bias on Ss' test performance. Data showed that spontaneity changes were independent of E 's perception. The hypnotic Ss gave fewer original responses in the conflict condition whether they were perceived as real or faking, and simulators likewise increased new responses independently of E 's perception. Similarly, E 's perception did not discriminate hypnotic or simulating Ss on the degree of complex relatedness given in their TAT stories. However, in Condition II, E perceived six of the seven Ss who identified the TAT figure according to their own sex as good hypnotic Ss. For the hypnotic Ss, then, identification with the ambiguous figure may be an index of their involvement under induced conflict, or may reflect E 's bias that S should identify with the figure in a particular way.

^a Actual frequency scores differed between the two raters, but the pattern of results reported here was identical for both raters.

^a Two raters scored the stories for complex relatedness. The interrater reliability for the measure was .78. Results are reported for the rater who, over all groups, rated more stories as complex-related.

DISCUSSION

The artificial induction of conflict had considerable effect on S's test behavior, but initial differences in the control testing between hypnotic and simulating Ss reduced differentiation of Ss in the conditions of induced conflict. There was evidence, though, that reduction of spontaneity was associated with hypnotic rather than simulating performance. Hypnotic Ss showed fewer unusual association responses in Conditions II and III while simulating Ss increased the number of original responses, and hypnotic Ss ceased giving unusual themes to the TAT card while simulating Ss increased the number of such themes reported. Both these findings suggest that fluidity of thought organization may have been impaired for the hypnotic Ss and that their mental processes were interfered with by the conflicting, emotionally laden ideas pertaining to the complex. The decrease in spontaneity suggests some degree of maladjustment since the nature of hypnotic Ss' responses to the test stimuli deviated from their usual style of responding. Results, however, are not completely unequivocal. It is conceivable that the act of simulating accounted for the increase in original responses simulators gave under the conflict conditions. If this is so, the simulating S's failure to respond in a similar fashion to the hypnotic S does not indicate that loss of spontaneity is a genuine effect of hypnosis, but, rather, the loss may be attributed to the psychological situation in which simulators were placed.

The main consensus of evidence was that simulators complying with the demands of the experimental situation displayed essentially the same pattern of behavior as did hypnotic Ss in the conflict situation. In the conflict condition, simulators demonstrated similar commonality scores, emotional themes to projective test material, and distortions in perception of semantic material, as did hypnotic Ss. The pattern of results suggests that previous findings on the effect hypnosis has on personality after the induction of conflict (e.g., Reyher, 1961) can be equally well explained by the demand characteristics of the experimental situation as by appeal to the effects of hypnotic treatment. This is not to say that

hypnosis failed to produce genuine effects, but the results of previous studies must be regarded as equivocal without adequate control for S expectations.

The extent to which test behavior of neurotically anxious Ss can be simulated draws attention to the need to develop clinical tests which are not amenable to conscious control. Present findings support the conclusion of Meltzoff (1965) that Ss can easily manipulate their responses on standard personality tests. The sentence completion test, for example, largely provides only information that S is "willing to give." This questions the situational validity of such tests. The problem is not so much one of estimating the extent to which such tests will behave in a predictable fashion following induced changes in the testee, but the extent to which S's test performance reflects the personality change and is not an artifact of the experimental situation.

There was some evidence that Condition III discriminated hypnotic and simulating Ss. In this condition the source incident was left intact but amnesia for it was removed. The semantic differential change scores for hypnotic Ss from Condition I to III bore no relationship to those of simulating Ss, and simulators reported fewer complex related stories to the TAT. Hypnotic Ss expressed greater subjective involvement, projecting feelings of personalized inadequacy into their TAT stories while simulators took refuge in the relatively "safe" category of exhaustion. These findings indicate that the demand characteristics were less obvious for this condition than for Condition II. Simulators apparently thought some difference in behavior was expected of them from one condition to the next. The evidence, in support of data from the postexperimental inquiry, indicates that some simulating Ss believed that the conflict should be alleviated when awareness of the source incident was present. The differences between the two groups in Condition III argue for the validity of the behavioral effects observed for hypnotic Ss in that condition.

Hypnotic Ss' behavior in Condition III indicated that repression was not a necessary condition for the occurrence of painful emotional states. Condition II and III were be-

haviorally indistinguishable and both demonstrated the presence of conflict. The removal of amnesia did not itself result in relief of the emotional experience. Neither did it result in greater conflict value as might be suggested by the theorizing of Freud, which states that repression protects *S* from painful emotional states, its removal meaning that the patient must "face pain." It is possible, however, as Bobbitt points out that the development of full awareness of conflict may accompany changes in disturbance so that maximum disturbance occurs when awareness is partial, but less than complete.

Hypnotic *Ss* characteristically expressed considerable subjective involvement under induced conflict, identifying with the ambiguously defined figure more than simulating *Ss*. The subjective reality of their feelings argues for changes and distortions in their behavior which accord with individual differences in the personality dynamics of *Ss*. In support of this, the evidence showed that hypnosis had no single effect on *Ss*' personalities even though the same paramnesia was suggested to all *Ss*. Subjective reactions ranged from expression of somatic symptoms such as headaches to symptoms of unspecified distress, paranoid ideas, and aggressive urges. Hypnotic *Ss* all reported that they experienced amnesia for the conflict material yet showed varying signs of breakdown of ego defenses. The most extreme reaction came from a hypnotic *S* who was agitated enough to hurl the TAT card at the *E* and say that he felt as if the police were watching him through a one-way screen which was placed in the room.

In conclusion, it should be stressed that, within the present methodology, null findings between groups can be unambiguously interpreted as indicating that *S* expectations could be a sufficient explanation for the results. Differences between groups are more difficult to interpret for the reason that simulators and hypnotic *Ss* are not equally susceptible to trance, and such differences could, presumably, be correlated with other personality and motivational traits which may complicate interpretation of performance. Interpretation of differences is also complicated by *E*'s instructions to *Ss* to simulate. Simulating instructions can be expected to yield specific

treatment effects which require independent study. Alternative methodologies such as those employed by Reyher (1961) and Moore (1964) which use only susceptible *Ss* in both the hypnotic and pretense conditions overcome the problem of personality differences among *Ss*, but fail to control for the demand characteristic effects resulting from using *Ss* as their own controls.

The limitations of the present design do not negate the main finding of the experiment: many of the signs of so-called maladjustment and emotional disturbance as inferred from *Ss*' behavior after the induction of conflict can be equally well explained by the expectations experimental *Ss* bring to their task, as by theories of hypnosis. Where differences do exist between simulating and hypnotic performance the evidence suggests that the anxiety-toned conflict has a repressive influence which reduces the spontaneity of the conflicted *Ss*. Results also show that repression is not a necessary condition for the occurrence of painful emotional states.

REFERENCES

- BOBBITT, R. A. The repression hypothesis studied in a situation of hypnotically induced conflict. *Journal of Abnormal and Social Psychology*, 1958, 56, 204-212.
- BRICKNER, R. M., & KUBIE, L. S. A miniature psychotic storm produced by a super-ego conflict over simple posthypnotic suggestion. *Psychoanalytic Quarterly*, 1936, 5, 467-487.
- COUNTS, R. M., & MENSE, I. N. Personality characteristics in hypnotically-induced hostility. *Journal of Clinical Psychology*, 1950, 6, 325-330.
- DAMASER, E. C., SHOR, R. E., & ORNE, M. T. Physiological effects during hypnotically requested emotions. *Psychosomatic Medicine*, 1963, 25, 334-343.
- DORLAND, W. A. *Illustrated medical dictionary*. Philadelphia: Saunders, 1965.
- ERICKSON, M. H. The method employed to formulate a complex story for the induction of an experimental neurosis in a hypnotic subject. *Journal of General Psychology*, 1944, 31, 67-84.
- ERON, L. D. A normative study of the Thematic Apperception Test. *Psychological Monographs*, 1950, 64(9, Whole No. 315).
- HUSTON, P. E., SHAKOW, D., & ERICKSON, M. H. A study of hypnotically induced complexes by means of the Luria technique. *Journal of General Psychology*, 1934, 11, 65-97.
- LEVITT, E. E., DEN BREKJEN, A., & PERSKY, H. The induction of clinical anxiety by means of a standardized hypnotic technique. *American Journal of Clinical Hypnosis*, 1960, 2, 206-214.

- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LURIA, A. H. *The nature of human conflicts*. New York: Grove Press, 1932.
- McDOUGALL, W. *Outline of abnormal psychology*. New York: Scribner's, 1926.
- MELTZOFF, J. The effect of mental set and item structure on response to a projective test. In B. I. Murstein (Ed.), *Handbook of projective techniques*. New York: Basic Books, 1965.
- MOORE, W. F. Effects of posthypnotic stimulation of hostility upon motivation. *American Journal of Clinical Hypnosis*, 1964, 7, 130-135.
- MURRAY, H. A. *Thematic Apperception Test*. Cambridge: Harvard University Press, 1943.
- ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.
- ORNE, M. T., & O'CONNELL, D. N. Diagnostic ratings of hypnotizability. *International Journal of Clinical and Experimental Hypnosis*, 1967, 15, 125-133.
- ORNE, M. T., SHEEHAN, P. W., & EVANS, F. J. Occurrence of posthypnotic behavior outside the experimental setting. *Journal of Personality and Social Psychology*, 1968, 9, 189-196.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- PALERMO, D. S., & JENKINS, J. J. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.
- REYHER, J. Posthypnotic stimulation of hypnotically induced conflict in relation to psychosomatic reactions and psychopathology. *Psychosomatic Medicine*, 1961, 23, 384-391.
- ROTTER, J. B., & RAFFERTY, J. E. *Manual: The Rotter Incomplete Sentences Blank, College Form*. New York: Psychological Corporation, 1950.
- SHOR, R. E., & ORNE, E. C. *Harvard Group Scale of Hypnotic Susceptibility, Form A*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- SWEETLAND, A. Hypnotic neuroses: Hypochondriasis and depression. *Journal of General Psychology*, 1948, 39, 91-105.
- WEITZENHOFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- WHITE, R. W. A preface to a theory of hypnotism. *Journal of Abnormal and Social Psychology*, 1941, 36, 477-505.

(Received June 29, 1967)

OBJECTIVE MEASURES OF SUSPICIOUSNESS¹

NOBLE A. ENDICOTT, SIDNEY JORTNER, AND EILEEN ABRAMOFF

Psychiatric Treatment Research Center, State University of New York, Brooklyn

The purpose of this study was to gather reliability and validity data on several measures of suspiciousness derived from psychological tests. Clinical ratings of suspiciousness were made on the basis of interviews with two groups of psychiatric patients: 90 hospitalized patients and 40 outpatients. Different "degrees" of suspiciousness were found to be characterized by qualitatively different types of Holtzman Inkblot Technique (HIT) content. Significant correlations were obtained in both groups of patients between the clinical ratings, the MMPI suspiciousness scale, and two content scales from the HIT. A multiple correlation combining the three test measures correlated .61 and .69 with rated suspiciousness in the hospitalized and outpatient groups, respectively.

The purpose of this study was to gather reliability and validity data on a number of measures of suspiciousness derived from psychological tests. It is hoped that studies of this type will eventually lead to the development of objective measures of the trait of suspiciousness which can be used in evaluating various forms of psychiatric treatment and in testing psychodynamic hypotheses.

Suspiciousness was defined in the present study as a psychological trait expressed by a distrust of others' motives in interpersonal relationships. This is closely related to, but not synonymous with, the diagnostic concept of paranoia. The present writers chose to study "suspiciousness" rather than "paranoia" because they thought psychological traits of this type would be more amenable to operational definition than the more global traditional diagnostic concepts.

A review of the literature failed to reveal any studies relating scores derived from psychological tests to suspiciousness, per se. However, there have been a number of previous investigations of the relationship between test scores and paranoia or paranoid tendencies (Bower, Testin, & Roberts, 1960; DuBrin, 1962; Friedman, 1957; Harris, 1949; Kleinmuntz, 1960; Murati, 1955; Nagge, 1951; Reznikoff & Nicholas, 1958; Ribler, 1957; Valentine & Robin, 1950; Wertheimer, 1953; Zeichner, 1956).

There is a general consensus among clinical writers that Rorschach content such as eyes,

masks, concealed or obscured figures, etc., frequently reflects paranoia or paranoid tendencies (Hertz, 1948; Lindner, 1947; Phillips & Smith, 1953; Schafer, 1954). The studies by Harris (1949) and DuBrin (1962) of Rorschach content generally supported these contentions, while Wertheimer's (1953) investigation of "eye" content failed to do so. In addition to the Rorschach, the TAT (Friedman, 1957; Nagge, 1951; Valentine & Robin, 1950; Zeichner, 1956), MMPI (Kleinmuntz, 1960), Figure Drawings (Reznikoff & Nicholas, 1958; Ribler, 1957) and the Aspiration Level Test (Murati, 1955) have been used to compare paranoids with nonparanoids.

The studies mentioned above suffer from a number of methodological limitations if one's goal is the development of an objective measure of suspiciousness. Most of these investigations compared the test scores of a paranoid group with those of one or more relatively nonparanoid groups. With this type of experimental design, it is not possible to determine whether the differences between the test scores of the groups are the consequence of differences between the groups in the "degree" of suspiciousness, or whether they are due to other personality characteristics in which the groups also differ. In addition, the majority of these investigations have used tests with many psychometric limitations (e.g., the Rorschach) or have used scores derived from only one psychological test.

The present study attempted to cope with these problems by relating each of the proposed test measures of suspiciousness to a more highly graduated criterion measure of

¹This study was supported in part by United States Public Health Service, National Institute of Mental Health Grant MH-05518.

suspiciousness (i.e., ratings of suspiciousness on a 5-point scale). In addition, each of the measures used in this study was correlated with ratings of 41 other symptomatic or personality variables. This permitted an analysis of the relative contribution made by each of these psychological states or traits to the production of the proposed test measures of suspiciousness.

In an effort to deal with some of the other limitations of previous research in this area, the Holtzman Inkblot Technique (HIT; Holtzman, Thorpe, Swartz, & Herron, 1961) was used instead of the Rorschach, and a relatively extensive battery of tests was administered to each patient.

METHOD

Subjects

Three groups of patients were involved in the present study: a hospitalized group ($n=90$), an outpatient group ($n=40$), and a mixed group of hospitalized patients and outpatients ($n=84$) used in a previous study (Endicott & Endicott, 1963). The hospitalized group represents all of the patients admitted to the Psychiatric Treatment Research Center, State University of New York, Brooklyn, between May 1963 and October 1964, who had either a verbal or performance Wechsler Adult Intelligence Scale (WAIS) IQ score of 80 or above and who remained in the Center long enough to be tested and interviewed. Of these patients, 63 were male and 27 were female. The patients were primarily of lower- or lower-middle-class backgrounds and had an average of 11.3 yr. of formal education (range 6–18 yr.). The mean full scale WAIS IQ was 101.9 (range 75–131), and their mean age was 33.7 yr. (range 17–58). The diagnoses for this group were: alcohol addiction, 47; schizophrenic reaction, 32; personality disorder (other than alcoholism), 9; psychotic depressive reaction, 1; and manic-depressive reaction, 1.

The outpatient group ($n=40$) was taken from the senior author's private practice. Of these patients, 18 were male and 22 were female. They had a mean age of 32 yr. (range 16–59). The group was predominately middle class in background and had a mean of 13 yr. of formal education (range 8–20). The diagnoses for this group were: alcohol addiction, 2; schizophrenic reaction, 9; personality disorder, 12; and psychoneurotic reaction, 17.

The third group of 84 patients were involved in only one aspect of the study. They provided the data for the initial derivation of the new MMPI suspiciousness scale used in this study. A description of this group of patients is presented in a previous publication (Endicott & Endicott, 1963).

Procedure for the Hospitalized Patients

Shortly after the patients were admitted to the hospital they were interviewed jointly by the first two authors. Following the interview, both made independent ratings of the patient's suspiciousness and each of the other psychological variables being investigated. All ratings were made without knowledge of the test results. The product-moment reliability coefficient for the suspiciousness ratings of these two raters was .92. Following the independent rating, the authors arrived at a consensus rating in conference. The consensus rating was used as the criterion measure in this study.

A battery of psychological tests consisting of Series A of the HIT, the WAIS, the MMPI, the Draw-a-Person Test, the Embedded Figures Test (EFT), and a portion of Cattell's O-A battery was administered after the initial interview. A more detailed account of the procedure followed can be found in a previous publication (Endicott & Jortner, 1966).

Procedure for the Outpatients

A similar procedure was followed with these patients except they were rated only by the senior author after five sessions and were given a less extensive battery of psychological tests.

Suspiciousness Rating Scale

A 5-point scale ranging from (1) "minimal" to (5) "extreme" suspiciousness was used in making the ratings. Each point was briefly defined in terms of the apparent pervasiveness and "intensity" of the trait, as follows.

1. Minimal suspiciousness.—The *S* would be described by observers (friends, associates, interviewers, etc.) as "trustful" or "not particularly cautious" in his dealings with other people. This should not be taken to mean that an individual rated in this category never distrusts anyone. Rather that his tendency to distrust others is primarily determined by the "realities" of the situation and he does not chronically expect to be taken advantage of or disappointed in his dealings with other people.

2. Mild suspiciousness.—The *Ss* classified here would frequently be described by others as "cautious" or "somewhat over-cautious" in their dealings with other people. Most observers, associates, etc., would not be likely to spontaneously label individuals classified here as "distrustful." The feelings of distrust would be relatively infrequent and/or of relatively low "intensity." Suspiciousness of this degree would not be considered pathological or particularly striking by most people in our culture.

3. Moderate suspiciousness.—With suspiciousness of this level there would be a tendency on the part of many of the *Ss*'s friends and associates (or interviewers) to spontaneously observe and report distrust of others as one of his characteristics.

ized group. The MMPI *Pa* scale *T* scores correlated .29; the *S18* scale, .44, the SCSS I, .58; and the SCSS II, -.23, with rated suspiciousness. When the HIT SP responses were added to the SCSS I and II, their correlations with rated suspiciousness were .52 and -.46, respectively. These data are presented in Table 1.

A multiple correlation of .69 was obtained in the cross-validation group between rated suspiciousness and the scores of the three best measures developed on the validation group: the MMPI *S18* scale, the HIT SCSS I plus SP responses, and the HIT SCSS II plus SP responses.

DISCUSSION

The results of the study were encouraging. Using three of the measures developed in this investigation, a multiple correlation of .61 was found with rated suspiciousness in the developmental sample. Applying these same measures to a markedly different cross-validation sample, a multiple correlation of .69 was obtained. While correlations of this magnitude leave considerable room for improvement, they are sufficiently high to warrant use of these measures in evaluating suspiciousness in groups of patients.

The finding of two relatively distinct types of "suspicious" HIT responses (SCSS I and SCSS II) was interpreted as indicating that different "degrees" of suspiciousness are characterized by qualitatively different types of responses to the HIT. This is similar to an earlier finding which related qualitative differences in HIT content to different "intensities" of depression (Endicott & Jortner, 1966).

The HIT "suspicious popular" (SP) responses were added to both the HIT SCSS I and II on the assumption that two of the more significant processes involved in the production of the SP responses had opposing directional effects and tended to cancel one another insofar as the correlations with rated suspiciousness were concerned. On the one hand, as the severity of suspiciousness increased there is a tendency to perceive a larger number of faces, masks, eyes, etc., on inkblot tests. This tendency is reflected most strikingly in responses to relatively unstruc-

tured inkblots. As the inkblots become more structured and their form more closely comes to resemble a face or mask, etc., another process, or group of processes, assumes greater significance. It has been demonstrated that the severely emotionally disturbed tend to view the world idiosyncratically and, as a consequence, produce fewer popular responses to inkblot tests than do "normals" (Holtzman, Thorpe, Swartz, & Herron, 1961). Because of the inverse relationship between suspiciousness and "mental health," the inkblot percepts of the mildly suspicious tend to resemble those of the general population (i.e., populars) more closely than do those of the markedly suspicious group. By relying on the "suspicious" content of the remainder of the HIT protocol the authors hoped to delineate which of these processes was most relevant in each patient's case.

The method of assigning the HIT SP responses to the HIT SCSS I and II was clearly more effective in the case of the SCSS II (Table 1). In the outpatient group the addition of the SP responses actually lowered the correlation between rated suspiciousness and the SCSS I scores from .58 to .52. Because of the greater number of SCSS I responses in the outpatients than in the hospitalized group, the cutoff points developed in the latter group were too low for the outpatients. As a consequence, a larger number of SP responses was added to the SCSS I scores of the outpatient moderately suspicious group than to those of the markedly suspicious and the correlation was lowered. The present contextual scoring system is regarded as preliminary and further

TABLE 1
PRODUCT-MOMENT CORRELATIONS BETWEEN
RATED SUSPICIOUSNESS AND TEST SCORES

Group	Holtzman Inkblot Technique				MMPI ^a	
	SCSS I	SCSS II	SCSS I +SP	SCSS II +SP	<i>S18</i>	<i>Pa</i>
Hospitalized patients ^b	.45*	-.38*	.46*	-.43*	.36*	.19
Outpatients ^c	.58*	-.23	.52*	-.46*	.44*	.29

Note.—Abbreviations: SCSS = Suspiciousness Content Scoring System; SP = suspicious popular responses.

^a The hospitalized patients' MMPI correlations are based on 84 tests.

^b *n* = 90.

^c *n* = 40.

* *p* < .01.

TABLE 2
MEAN TEST SCORES AND RATINGS OF "DEGREE" OF SUSPICIOUSNESS

Clinical ratings of suspiciousness	HIT SCSS I		HIT SCSS II		HIT SCSS I+SP		HIT SCSS II+SP		MMPI <i>Si8</i>		MMPI <i>Pa</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hospitalized group (<i>n</i> 90) ^a												
Extreme (<i>n</i> = 4)	8.75		0.50		9.25		0.50		3.3		57.0	
Marked (<i>n</i> = 25)	5.52	4.91	0.35	0.49	5.88	5.30	0.35	0.49	6.7	3.4	67.8	14.0
Moderate (<i>n</i> = 23)	4.04	3.39	0.48	0.66	4.35	3.73	0.52	0.66	4.3	3.3	74.8	9.4
Mild (<i>n</i> = 38)	1.89	1.89	1.11	0.92	2.00	2.22	1.32	1.10	2.9	2.4	59.7	10.7
Total group (<i>n</i> = 90)	3.76	4.07	0.71	0.81	4.03	4.44	0.82	0.92	4.3	3.4	63.8	12.2
Outpatient group (<i>n</i> 40)												
Marked (<i>n</i> = 10)	9.70	4.24	0.20	0.42	10.40	4.84	0.20	0.42	5.1	3.5	66.6	11.7
Moderate (<i>n</i> = 22)	6.05	3.42	0.64	0.73	6.95	4.37	0.68	0.84	2.4	2.1	63.9	15.6
Mild (<i>n</i> = 8)	2.50	1.69	1.00	1.07	3.00	2.45	1.75	1.75	1.8	1.4	54.5	6.4
Total group (<i>n</i> = 40)	6.25	4.11	0.60	0.78	7.02	4.80	0.78	1.12	2.9	2.7	62.7	13.7

Note.—Abbreviations: HIT SCSS = Holtzman Inkblot Technique, Suspiciousness Content Scoring System; SP = suspicious popular responses.

^a For the MMPI scores the number of patients is: extreme, 4; marked, 24; moderate, 21; and mild, 35.

research is necessary in order to develop more sophisticated scoring systems of this type.

The finding of a higher correlation between rated suspiciousness and the HIT SCSS I scores in the cross-validation group (the outpatients) than in the hospitalized group is probably due in large part to the greater number of SCSS I responses in the former group (Table 2). The reasons for this difference between the two groups is unknown. The SCSS I scores were not significantly correlated with WAIS IQ, education, or field dependence-independence as measured by the EFT (Witkin, Dyk, Faterston, Goodenough, & Karp, 1962). Whatever the causes of this difference, it is not a general finding with respect to all HIT content measures of "psychopathology." For example, the two groups did not differ significantly in HIT "depressive" content, but the hospitalized group had significantly more HIT content associated with somatic preoccupation than did the outpatient group (Endicott & Jortner, 1966, 1967).

Another factor which may have contributed to the difference between the correlations of the ratings and the test scores for the two groups was a lower accuracy of the ratings in the hospitalized group. Suspiciousness is

generally considered a socially undesirable trait and is frequently denied. In the hospitalized group the ratings were made after one interview of 60-90 min. duration. On several occasions it was later discovered, through conversations with the patient's psychotherapist, nurses, or other patients, that the authors had underestimated the degree of the patient's suspiciousness. This occurred most frequently among the alcoholic patients.

In the outpatient group the ratings were usually made after five 1-hr. interviews, and as a consequence the rater had considerably more information about them than about the hospitalized group. In addition, the outpatients came to the interviews under conditions more conducive to self-revelation than did the hospitalized patients.

The findings of the present study suggest a number of procedures whereby more valid objective measures of suspiciousness might be developed. For example, some of the scoring categories in the SCSS I were consistently more discriminating than others. If the same differences in discriminatory power are found in other groups, a weighting system could be developed to use this differential.

The finding that different "degrees" of suspiciousness are characterized by qualita-

tively different types of HIT responses suggests another method of developing a more valid method of evaluation. For example, using the combined patient group of 130, there were 12 patients who had the pattern SCSS I = 0 and SCSS II = 0-3. Ten of these patients were rated mildly suspicious and two were rated moderately suspicious. Several other HIT score patterns had a similar discriminatory power. By using such highly discriminatory HIT patterns, and other measures derived from tests, in a method based on the decision tree model (Kleinmuntz, 1963), it should be possible to develop objective evaluative techniques which are much more reliable and valid than the relatively simple linear methods discussed in this report.

REFERENCES

- BOWER, P. A., TESTIN, R., & ROBERTS, A. Rorschach diagnosis by a systematic combining of content, thought process, and determinant scales. *Genetic Psychology Monographs*, 1960, 62, 105-183.
- DUBRIN, A. J. The Rorschach "eyes" hypothesis and paranoid schizophrenia. *Journal of Clinical Psychology*, 1962, 18, 468-471.
- ENDICOTT, N. A., & ENDICOTT, J. Objective measures of somatic preoccupation. *Journal of Nervous and Mental Disease*, 1963, 137, 427-437.
- ENDICOTT, N. A., & JORTNER, S. Objective measures of depression. *Archives of General Psychiatry*, 1966, 15, 249-255.
- ENDICOTT, N. A., & JORTNER, S. Correlates of somatic concern derived from psychological tests. *Journal of Nervous and Mental Disease*, 1967, 144, 133-138.
- FRIEDMAN, I. Characteristics of the Thematic Apperception Test heroes of normal, psychoneurotic, and paranoid schizophrenic subjects. *Journal of Projective Techniques*, 1957, 21, 372-376.
- HARRIS, W. E. A study of Rorschach factors in paranoid schizophrenia. Unpublished doctoral dissertation, University of Kentucky, 1949.
- HERTZ, M. Suicidal configurations in Rorschach records. *Rorschach Research Exchange and Journal of Projective Techniques*, 1948, 12, 3-58.
- HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality*. Austin: University of Texas Press, 1961.
- KLEINMUNTZ, B. Two types of paranoid schizophrenia. *Journal of Clinical Psychology*, 1960, 16, 310-312.
- KLEINMUNTZ, B. MMPI decision rules for the identification of college maladjustment: A digital computer approach. *Psychological Monographs*, 1963, 74(14, Whole No. 577).
- LINDNER, R. M. Analysis of the Rorschach Test by content. *Journal of Clinical Psychopathology*, 1947, 8, 707-719.
- MURATI, J. D. Differences in test behavior between paranoid and non-paranoid schizophrenic groups. Unpublished doctoral dissertation, University of Wisconsin, 1955.
- NAGGE, W. W. A study of the behavior of paranoid schizophrenics on the Thematic Apperception Test. Unpublished doctoral dissertation, University of Kentucky, 1951.
- PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: Advanced technique*. New York: Grune & Stratton, 1953.
- REZENIKOFF, M., & NICHOLAS, A. L. An evaluation of human figure drawing indications of paranoid pathology. *Journal of Consulting Psychology*, 1958, 395-397.
- RIBLER, R. I. Diagnostic prediction from emphasis on the eye and ear in human figure drawings. *Journal of Consulting Psychology*, 1957, 21, 223-225.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing: Theory and application*. New York: Grune & Stratton, 1954.
- VALENTINE, M., & ROBIN, A. A. Aspects of Thematic Apperception Testing: Paranoid schizophrenia. *Journal of Mental Science*, 1950, 96, 869-888.
- WERTHEIMER, M. On the supposed behavioral correlates of an "eye" content response on the Rorschach. *Journal of Consulting Psychology*, 1953, 17, 189-194.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- ZEICHNER, A. Conception of masculine and feminine roles in paranoid schizophrenia. *Journal of Projective Techniques*, 1956, 20, 348-354.

(Received July 5, 1967)

DOMINANCE AND CONFLICT IN THE INTERACTIONS BETWEEN PARENTS OF NORMAL AND NEUROTIC CHILDREN¹

SUZANNE GASSNER

AND

EDWARD J. MURRAY

Syracuse University

University of Miami

The Revealed Differences Technique was used to study dominance and conflict patterns in the interaction of neurotic and normal children of both sexes. Two measures of relative dominance indicated that neurotic boys have dominant mothers and neurotic girls have dominant fathers. Two measures of conflict indicated a more hostile interaction between the parents of neurotic children. In the neurotic group, there was a relationship between the areas of husband-wife conflict and independent clinic ratings of corresponding problem areas in the behavior of the child. The results were interpreted in terms of an identification learning conflict.

In recent years, there has been a considerable focusing of interest on the family situation as a major determinant of a wide variety of neurotic, delinquent, and psychotic patterns of pathological behavior (Ackerman, 1958; Lidz, Fleck, & Cornelison, 1965; Spiegel & Bell, 1959). It has been suggested that the development of pathological behavior in the child is related not only to emotional disturbances in the individual parents but also to disturbances in the relationship between parents which may affect the mode of functioning of the entire family. Along with this, a major movement to develop and apply psychotherapeutic techniques to family groups has emerged (e.g., Boszormenyi-Nagy & Framo, 1965; Satir, 1964). Unfortunately the empirical base for these theoretical and therapeutic developments is exceedingly weak. Most of the work has consisted of clinical observations, and the more systematic clinical research has been characterized by a lack of

comparison groups, the use of inadequate measuring procedures, and a failure to control for variables such as age, sex, social class, and type of pathology (Fontana, 1966; Frank, 1965).

One of the major problems in this area is how to reconstruct the family situation as it may have been during the presumably critical early years of the childhood of the disturbed individual. Retrospective accounts by either the individual or his parents are subject to a number of distortions and measurement difficulties. Therefore, attention has been turned toward a direct study of the current personalities in the family and their interaction with one another. Fontana (1966) has cogently argued for the utility of this method in the absence of longitudinal studies, while at the same time pointing out its limitations.

The direct study of individual members of a family with a pathological child suggests that the parents are maladjusted in various ways (Fisher, Boyd, Walker, & Sheen, 1959; Peterson, Becker, Hellmer, Shoemaker, & Quay, 1959; Winder, 1960). However, it has been difficult to specify these parental characteristics, particularly in relation to different types of pathology. Many investigators now believe that the interaction between family members is a more significant factor than individual personalities (Handel, 1965). The most widely used measure of parental interaction is the Revealed Differences Technique (RDT; Strodbeck, 1951), an experimental procedure in which parents answer standardized questions individually and are then brought to-

¹ This study is based on a master's thesis at Syracuse University submitted by the senior author. The junior author supervised the thesis and assumed responsibility for the preparation of the final report. The authors wish to express their appreciation to the staffs of the Onondaga County Child Guidance Center and the Liverpool, New York, School System for their cooperation in obtaining families. The authors also wish to acknowledge the particularly helpful participation of Michael Boucher, Robert Deysack, Jeffrey Elenewski, Sally Sanford, and David Streiner as well as that of Gordon Carr, Marilyn Davis, Martin Gutride, Linda Lee, John Lemerond, John McManus, and Sandra Siegel. Finally, the authors thank David Hunt, Donald Meyers, Conrad Schwartz, and Paul Wohlford for their helpful comments and suggestions.

gether to reconcile their views. Unfortunately, nearly all of the studies using the RDT have been concerned with schizophrenia (e.g., Farina, 1960).

The purpose of the present study is to examine, with the aid of the RDT and with better controlled conditions, the interaction of parents of neurotic children in comparison with the interaction of parents of normal children. The interest in neurosis was determined by the relative neglect of this level of psychopathology in the family research literature as compared with the relative importance of neurotic phenomena in the formulation of classic and current theories of personality and psychopathology.

In addition, the use of a child guidance center population of neurotic children has several advantages in the general study of family psychopathology. First, the study of families is more feasible in this situation than many others. Second, the use of this young outpatient population with relatively acute problems minimizes the influence of chronicity and hospitalization that complicates the interpretation of many studies using adult psychiatric patients. Third, the close temporal proximity of the child's pathology and the family patterns studied, while not eliminating the problem of interpreting causal effect, makes such interpretations a little more credible than when one must extrapolate back 20 yr. or more.

The first variable studied was parental dominance. This has been considered a critical variable in psychopathology since the early work of Levy (1945) but the supporting evidence is rather weak (Fontana, 1966; Frank, 1965). There are actually several versions of the dominance hypothesis: first, that excessive maternal dominance (or alternatively paternal dominance) is a critical factor in pathological behavior regardless of the sex of the child; second, that an extreme disparity in the dominance of the parents, regardless of direction, produces a pathological situation for all children; third, there is a variation emphasizing the interaction between the sex of the dominant parent and the sex of the disturbed child. All of these versions assume that abnormal dominance patterns interfere with the sex-role-identification learning of the

child. These three variations of the dominance hypothesis were tested with measures from the RDT using parents of normal and neurotic children.

The next variable studied was conflict between the parents. It has been suggested that neurotic behavior patterns could result from a child being drawn into a conflict between the parents (Vogel & Bell, 1960). However, Fisher et al. (1959) found little conflict with parents of adult neurotics as compared to parents of schizophrenics. Parental conflict has been found rather consistently with schizophrenia (Fontana, 1966). In the present study, the hypothesis that parents of neurotic children show greater conflict than parents of normal children was tested with measures from the RDT.

The conflict theory, of course, supposes not only that the parents be in conflict but that the child's behavior is affected by this parental conflict. The objective evidence for this is extremely limited. Peterson et al. (1959) and Becker, Peterson, Hellmer, Shoemaker, and Quay (1959) did find some relationship between parental characteristics and the problems presented by emotionally disturbed children at a child guidance clinic. Unfortunately, the child's problem was assessed from the same parental interview on which the parent's personality was judged. Therefore, the present study incorporated a method of assessing three areas of conflict between parents—achievement, affiliation, and aggression—which could be related to independent ratings of three areas of the neurotic child's pathology—underachieving, social withdrawal, and aggressive behavior. It was hypothesized that there would be a positive relationship between the areas of parental conflict and the neurotic child's areas of difficulty.

If a relationship were found between areas of parental conflict and the child's problem, there still would be no way of identifying the antecedent variable. The child's problem could have grown out of the parental problem or the parental problem might simply represent a reaction to the child's problem. Although the problem of causal direction will not be solved for a long time in any definitive way, it was thought that some relevant information might be obtained. Therefore, the

parents were asked two types of questions—half about parent-child situations and half about husband-wife situations having nothing to do with children. It was assumed that if the parental conflicts merely reflected a response to the child's problem, the relationship between area of child difficulty and area of parental conflict should be greater with the parent-child situations. On the other hand, if the child's behavior was in response to a parental problem, the relationship should be as strong or stronger with husband-wife situations.

METHOD

Subjects

The *Ss* consisted of the *parents* of 30 neurotic children, 18 male and 12 female, at the local child guidance clinic and the *parents* of 30 normal children, 16 male and 14 female, obtained through a local school system. Parents of children with organic, psychotic, or legally delinquent problems were excluded from the neurotic group. Other than this, successive referrals were seen as part of the diagnostic procedure with no refusals. The children were typical of neurotic-level child guidance cases with school problems, anxiety symptoms, and social maladjustment. Parents of children with any psychopathology known to the school psychologist were excluded from the normal group. Other than this the normal families were selected at random by computer and then asked to participate in a study of normal families. About 45% of the normal families contacted agreed to take part in the study. Since Zuckerman, Barrett, and Bragiel (1960) found that the parents of the most maladjusted school children were the most reluctant to participate in this sort of study, the voluntary participants in the present study presumably represent a well-adjusted sample.

The families in the two groups were not significantly different with respect to the ages of the father, mother, or child. The ages of the children in both groups ranged from 6 to 16, with a mean of about 11. The social class of the families was measured by the Warner (1960) scales for father's occupation and income with good reliability between two independent judges ($r = .84$, $p < .01$). Both groups consisted predominately of lower-middle-class families and were not significantly different on this variable.

Procedure

The parents in the neurotic group were seen at the child guidance clinic and the parents in the normal group at the university by the same female graduate student. All parents were interviewed first individually and then together. In the individual interview each parent was asked what he or she would do in 12 hypothetical situations. In the joint interview the same situations were read and the

parents were asked to come to an agreement on how they would react. The instructions closely followed those used by Farina (1960).

The 12 situations were designed to touch upon three motivational systems—4 were about achievement, 4 were about affiliation, and 4 were about aggression. Four graduate students in clinical psychology each correctly identified the motive involved in well over 90% of the situations. In addition, half of the situations, two from each motivational group, dealt with a problem between a husband and wife and the other half with a problem between a parent and a child. In half of the husband-wife situations, the husband was the central character, and the wife in the other half. The child was the central character in the parent-child situations. The situations were presented, for all *Ss*, in a fixed order that was randomly determined except that successive situations did not touch upon the same motive. While these hypothetical situations were newly created, they are of the same genre as those used by Farina. An example is: "A wife very much wants to live in a better neighborhood. The only way the family could afford this is if she takes a part-time job. If you were the [wife, husband], what would you think she should do?" This is a husband-wife situation, with the wife as the main character, dealing with achievement motivation.

In order to minimize *E* bias, *E* was provided with a fixed program of appropriate responses (e.g., "um-hmm," "o.k.") for use after each answer of the *S*. In addition, *E* carefully controlled facial expressions and glances. As a further check on the main *E*, who saw 20 sets of parents in each group, seven graduate students in clinical psychology, who were unaware of the specific purpose of the study or the nature of the measures to be used, interviewed the other 10 sets of parents in each group. Later, these were compared with 10 randomly selected sets of parents from each group seen by the main *E* and no significant, or even suggestive, differences were found on any of the measures to be described below.

Measures

The analysis was based on the taped interview with the parents together. The measures, taken directly from Farina (1960), concerned *dominance* and *conflict*.

The measures for *dominance* were: (a) *Speaks First and Last* (number of times the father spoke first or last during the interview minus the number of times the mother did the same); (b) *Passive Acceptance* (the number of the 12 situations in which the mother simply agreed with the other's solution minus the number of times the father behaved similarly); (c) *Total Time Spoken* (difference in time in seconds during which father and mother spoke). Positive scores indicated paternal dominance and negative scores maternal dominance. Dominance measures could be expressed either as relative dominance of one parent over the other or as absolute disparity in dominance. The measures could also be related to sex of child.

The measures for Conflict were (a) Disagreements and Aggressions (the number during the interview by father plus mother); (b) Failure to Agree (the number of the 12 situations with no mutually acceptable solution); (c) Simultaneous Speech (length of time in seconds during which both parents spoke concurrently); (d) Interruptions (the number of times either parent interrupted the other); (e) Verbal Activity (total amount of time spoken by father and mother).

The tapes were scored blind by the main *E* approximately 6 mo. after the data were gathered. As a reliability check, two graduate students in clinical psychology, unaware of the purpose of the study, scored 10 randomly selected tapes from each of the groups. The scores were quite similar and rank-order correlations between the main *E* and the naive judges ranged from .79 to .99 ($p < .01$ or better) for the various measures.

Ratings of Neurotic Children

Preliminary discussions with the clinic staff suggested that the neurotic children could not be easily grouped into those with only achievement, affiliation, or aggression problems since these behavioral features frequently overlapped. Therefore, each child was rated on all three variables. The staff member of the child guidance clinic most familiar with the child was asked to make the ratings at the end of the intensive 3-hr. diagnostic procedure. Three 7-point scales ranging from "very little problem" to "great difficulty" were used to rate the problem areas of underachieving, social withdrawal, and aggressive behavior. The nature of the diagnostic procedure precluded an adequate assessment of the reliability of these scales.

RESULTS

Parental Dominance Hypotheses

The first version of the dominance hypothesis tested was that the parents in the neurotic group as a whole would show a pattern of maternal dominance, or alternatively a pattern of paternal dominance, as contrasted with the parents in the normal group. Table 1 shows the mean relative dominance scores for

TABLE 1

MEAN RELATIVE DOMINANCE MEASURES FOR PARENTS OF NEUROTIC AND NORMAL CHILDREN

Measure	Parents of neurotics	Parents of normals	F
Speaks First and Last (No. father minus no. mother)	1.33	-.60	.953
Passive Acceptance (No. mother minus no. father)	-.17	-.53	.814
Total Time Spoken (Sec. father minus sec. mother)	71.13	18.27	1.954

TABLE 2

MEAN ABSOLUTE DOMINANCE MEASURES FOR PARENTS OF NEUROTIC AND NORMAL CHILDREN

Measure	Parents of neurotics	Parents of normals	F
Speaks First and Last (Father-mother difference)	6.33	4.60	1.870
Passive Acceptance (Father-mother difference)	1.03	1.00	.009
Total Time Spoken (Father-mother difference)	128.60	86.73	2.286

the three dominance measures for the parents in each group. If the neurotic parents as a group showed maternal dominance the scores should be relatively more negative than those of the neurotic group. If anything, the neurotic parents are slightly more paternal in dominance patterns. Actually, none of these differences is statistically significant. The data in this section were evaluated statistically by a computer program for multiple analysis of variance appropriate for groups of equal or unequal size developed by Clyde, Cramer, and Sherin (1966). Thus, the parents of the neurotic group as a whole do not appear to show excessive maternal or paternal dominance.

Next, the idea that parents of neurotic children show an extreme disparity in the dominance of the parents, regardless of direction, was tested by examining the absolute differences in the three dominance measures. The results are shown in Table 2. Although there is a trend for parents of neurotic children to show a greater dominance discrepancy than parents of normal children, none of these differences is statistically significant. Therefore, dominance discrepancy between the parents does not appear to be a strong variable in the family backgrounds of these neurotic children.

Finally, the version of the dominance hypothesis emphasizing the interaction between the sex of the dominant parent and the sex of the neurotic child was examined. The mean relative dominance measures for parents of boys and girls in the neurotic and normal groups are shown in Table 3. The Speaks First and Last and the Passive Acceptance measures show a significant interaction between group and sex of the child. Both measures sug-

TABLE 3

MEAN RELATIVE DOMINANCE MEASURES FOR PARENTS OF NEUROTIC AND NORMAL BOYS AND GIRLS

Measure	Parents of neurotic		Parents of normal		<i>F</i>
	Boys	Girls	Boys	Girls	
Speaks First and Last	-2.22	6.17	.00	-1.29	7.256*
Passive Acceptance	-.67	.58	-.06	-1.07	7.547*
Total Time Spoken	68.83	74.58	2.88	35.86	.127

* $p < .01$.

gest that neurotic boys come from maternally dominated families and the girls from paternally dominated families. The measures for the parents of normal boys do not show much deviation in either direction while the parents of normal girls are slightly in the maternal direction. The Total Time Spoken measure does not show a statistically significant interaction between group and sex of child, possibly because of the tendency of fathers in all groups to talk more.

Analyzing the data in another way, there is a significant difference between the parents of neurotic boys and girls on Speaks First and Last ($F = 8.721$, $p < .01$) and Passive Acceptance ($F = 4.571$, $p < .05$). The parents of normal boys and girls do not differ significantly on these measures. So, too, the parents of neurotic girls and the parents of normal girls are significantly different on Speaks First and Last ($F = 8.540$, $p < .01$) and Passive Acceptance ($F = 5.318$, $p < .05$). The corresponding differences for parents of neurotic and normal boys do not reach statistical significance, although they are in the same direction. Thus, the results are most striking for the parents of neurotic girls.

Do the averaged data reflect a few extreme cases or a more prevalent association between parental dominance and sex of neurotic child? Using the Speaks First and Last measure, each family was simply classified as showing paternal or maternal dominance on the basis of the sign of the score with ties eliminated. These frequency data are shown in Table 4 where it can be seen that a substantial proportion of the families of neurotic boys were classified as maternal and an even larger proportion of the families of neurotic girls were classified as paternal (corrected $\chi^2 = 5.1$,

TABLE 4

FREQUENCY OF PATERNAL OR MATERNAL DOMINANCE IN PARENTS OF NEUROTIC AND NORMAL BOYS AND GIRLS

Dominance based on speaks first or last measure	Parents of neurotic		Parents of normal	
	Boys	Girls	Boys	Girls
Paternal dominance	4	9	7	5
Maternal dominance	10	2	6	6

$df = 1$, $p < .05$). The normal boys and girls appear to come about equally from paternal and maternal homes (corrected $\chi^2 < 1$, $df = 1$, ns). The Passive Acceptance measure shows a similar trend but, because of a larger number of ties, the results are not as striking.

An examination of the individual data in the four subgroups suggests that the groups are roughly comparable in the size of discrepancies between the dominance of father and mother. While the two or three most extreme discrepancies occurred in the neurotic groups, there are no significant differences between the subgroups using the three absolute dominance measures nor did the absolute dominance measures show significant interactions between group and sex of child. It would be more accurate to say that neurotic children more frequently came from homes showing cross-sex parental dominance than that neurotic children came from homes showing a greater amount of cross-sex parental dominance.

Parental Conflict Hypothesis

The results for the five measures of conflict are presented in Table 5. The data were evaluated statistically by the computer program for multiple analysis of variance developed by

TABLE 5

MEAN CONFLICT MEASURES FOR PARENTS OF NEUROTIC AND NORMAL CHILDREN

Measure	Parents of neurotics	Parents of normals	<i>F</i>
Disagreements and Aggressions (no.)	4.77	.63	6.975*
Failure to Agree (no.)	.83	.17	6.947*
Simultaneous Speech (sec.)	6.00	6.30	.031
Interruptions (no.)	12.83	16.07	1.216
Verbal Activity (sec.)	488.07	378.33	2.137

* $p < .02$.

Clyde et al. (1966). There are significantly more instances of Disagreements and Aggressions between the parents of neurotic children than between the parents of normal children. The parents in the neurotic group failed to reach a mutually acceptable solution in significantly more situations than did the parents in the normal group. The other three measures did not differentiate the two groups at a statistically significant level. The pattern of results was similar for parents of boys and girls, but statistically stronger for the boys. There was no significant interaction between group and sex of child on any of the conflict measures.

Qualitatively, the groups appear to be differentiated on the basis of overt parental hostility. The impression gained from listening to the tapes was that the first two measures reflected a hostile interaction. It is interesting to note that these are the same two measures judged most valid by Farina and Dunham (1963). The other three measures, however, seemed to represent a positive give-and-take discussion as often as a hostile interaction. Farina and Dunham, as well as Hetherington and Frankie (1967), combined several of these conflict indicators to get an overall conflict measure, but this did not appear to the present authors to be legitimate because the indicators are not all independent of one another.

The hostile interaction in the neurotic group was about evenly initiated by the father and the mother. The parents were very close on the measures of Disagreements and Aggressions and Interruptions. There was a non-significant trend for the father to talk more. The Failure to Agree and Simultaneous Speech measures could not be broken down in this way. Furthermore, the data did not suggest an interaction between sex of parent and sex of child on any of the three relevant measures of conflict.

Relationship between Parental Dominance and Conflict

The results so far show that while cross-sex parental dominance is more frequent in the neurotic group it also occurs in a substantial proportion of the normal families. The groups are also differentiated on the basis of

interparental conflict. The question arises as to whether the cross-sex parental dominance in the neurotic group is accompanied by interparental conflict while in the normal group cross-sex parental dominance occurs without appreciable interparental conflict. In order to answer this question, both neurotic and normal families were classified as showing cross-sex or same-sex parental dominance on the basis of the Speaks First and Last measure as was done for Table 4. They were also classified as high conflict if they had two or more Disagreements and Aggressions or as low conflict if they had one or less.

The results of this analysis are shown in Table 6 where it can be seen that for neurotic families most of the cross-sex parental dominance cases also show high conflict but in the normal families cross-sex parental dominance occurs without high conflict (corrected $\chi^2 = 9.2$, $df = 1$, $p < .01$). When the boys with dominant mothers are considered separately, the same relationship is found. Because of small expected frequencies this relationship was tested with the Fisher exact probability test (Latscha, 1953; Siegel, 1956) and found to be significant ($p < .01$). The results for girls with dominant fathers were in the same direction but failed to reach significance (Fisher test).

The data for children with the same-sex parent dominant are shown in the right half of Table 6 for comparison. The neurotic and normal groups are not significantly different (Fisher test). In comparing the neurotic children with same-sex and cross-sex parental dominance, there is a trend for the former to show more frequent low-conflict cases as opposed to the high-conflict cases in the latter, but this reversal just failed to reach statistical significance.

TABLE 6
PARENTAL DOMINANCE AND PARENTAL CONFLICT

Parents of	Cross-sex parent dominant		Same-sex parent dominant	
	High conflict	Low conflict	High conflict	Low conflict
Neurotic children	14	5	2	4
Normal children	1	10	1	12

Note.—Frequency of high and low interparental conflict in families with cross-sex and same-sex parental dominance for neurotic and normal children.

cal significance (Fisher test). A comparison of the normal families in the same-sex and cross-sex parental dominance groups shows no significant difference in frequency of high- and low-conflict cases (Fisher test).

Relationship between Parental Conflict and Problems of Neurotic Child

As anticipated, most of the neurotic children did have a variety of combinations of problems of underachievement, social withdrawal, and aggressive behavior according to the ratings of the clinic staff. These problems were about equally represented in the group of neurotic boys but there were fewer neurotic girls with a problem of aggressive behavior. In general, the procedure of rating degree of difficulty in several areas provided a much more meaningful picture of the neurotic child's problems than a simple grouping, according to the most salient problem, would have.

It was predicted that there would be a relationship among the measures of parental conflict in the three areas and the degree of disturbance of the neurotic child in these same areas. Two technical problems arose in testing this relationship. First, when the various measures of parental interaction were broken down by content area, the values became quite small, with a preponderance of zeros. Second, in making ratings of the problem areas of the neurotic children, the staff members of the child guidance center tended to use quite different ranges of the rating scales making direct comparisons specious. These technical problems precluded the more obvious methods of analysis.

The relationship between parental interaction and child behavior was tested in somewhat an unusual fashion. First, one of the parental measures of conflict, Verbal Activity (total time spoken by father and mother), was not grossly distorted by the breakdown and showed an adequate variation among content areas. Thus, it was possible to rank, for each pair of parents, the amount of time spent in discussing the situations in the areas of affiliation, achievement, and aggression. Furthermore, although the center staff used different ranges of the scale, it was possible to rank the ratings of degree of difficulty in

the areas of withdrawal, underachieving, and aggressive behavior for each neurotic child. Thus, it was feasible to compute a rank-order correlation for each family between the Verbal Activity of the parents in three areas and the degree of difficulty of the child in the corresponding three areas.

In order to express the relationship between parental interaction and child behavior for the neurotic group as a whole, note should be taken of the fact that with an N of 3 only four values of ρ can be obtained: +1.00, +.50, -.50, and -1.00. The number of ρ s that would be expected by chance alone can be obtained from the normal probability table. The expected frequencies for 28 families (2 families were eliminated because of ties in the staff ratings) are shown in the top row of Table 7. The frequencies of ρ s obtained are shown in the second row of Table 7. It can be seen that there are many more positive rank-order correlations obtained than would be expected by chance alone ($\chi^2 = 9.2$, $df = 3$, $p < .05$). Therefore, there does seem to be a positive relationship between the relative length of time the parents discuss situations in the three areas and the relative degree of difficulty the child is seen as having in the corresponding areas.

The degree to which the problems of the child were related to parent-child versus husband-wife conflict was examined in the following way. Separate rank-order correlations were computed between the rankings of the child's difficulty in the three areas when only the two situations in each area dealing with husband-wife problems were included and when only the two situations in each area dealing with parent-child problems were in-

TABLE 7
PARENTAL CONFLICT AND PROBLEMS
OF NEUROTIC CHILD

Area	ρ			
	+1.00	+.50	-.50	-1.00
Chance expectation	4.3	9.5	9.5	4.3
Obtained 12 situations	9	12	6	1

Note.—Expected and obtained distributions of rank-order correlations between the length of time the parents discussed situations in the areas of affiliation, achievement, and aggression and the degree of difficulty the neurotic child was judged to have in these areas.

TABLE 8

HUSBAND-WIFE VERSUS PARENT-CHILD CONFLICTS

Area	rho			
	+1.00	+ .50	-.50	-1.00
Chance expectation	4.3	9.5	9.5	4.3
Obtained for six husband-wife situations	6	14	8	0
Obtained for six parent-child situations	5	12	5	6

Note.—Rank-order correlations between the length of time the parents discussed situations in the areas of affiliation, achievement, and aggression and the degree of difficulty the child was judged to have in these areas—distributions for chance expectation, obtained for husband-wife situations, and obtained for parent-child situations.

cluded. The results are shown in Table 8. The distribution of rhos for the husband-wife situations deviates in positive direction and is significantly different from the chance distribution ($\chi^2 = 7.89$, $df = 3$, $p < .05$). The distribution for the parent-child situations is only slightly in the positive direction and does not deviate significantly from chance ($\chi^2 = 3.59$, ns). Thus, the relationship between parental interaction and child behavior holds when the parents are discussing husband-wife situations rather than when they are discussing parent-child situations.

DISCUSSION

Before interpreting the results, several limitations of the study should be noted. First, the results should not be generalized beyond a population of lower-middle-class parents of school-age children with neurotic-level problems who are being seen at a child guidance center. Second, the parents in one group were seen as part of an intake procedure at a child guidance center while those in the other group were told that they were part of a study of normal families. While it is unlikely that these differing circumstances account for all of the results, their influence on the behavior of the parents could not be evaluated in this study. Third, while the critical measures from the Revealed Differences Technique (RDT) seem reasonably well validated in other studies, the frankly experimental ratings of the problems of the neurotic child are of unknown reliability and validity.

The results from the dominance measures

clearly support the hypothesis that neurotic behavior is related to the sex of the dominant parent and the sex of the disturbed child. In the present sample, neurotic behavior is not associated simply with dominant mothers nor simply with dominant fathers without respect to the sex of the child. So, too, neurotic behavior is not associated with a discrepancy in the dominance of the parents regardless of direction. Rather, neurotic boys tend to come from maternally dominated homes and neurotic girls from paternally dominated homes. Presumably, these situations produce difficulties in identification learning.

Nevertheless, an explanation of neurotic behavior on the basis of cross-sex parental dominance is inadequate. Recall that about half of the normal children come from homes showing about the same degree of cross-sex parental dominance. Obviously some other variable must be involved. The results from the conflict measures suggest that this additional variable is an overtly hostile conflict between the parents. One might be tempted to say that interparental conflict is a sufficient explanation of the neurotic patterns in this sample but this would ignore the intriguing results on the distribution of cross-sex parental dominance between the groups as well as the association of cross-sex parental dominance and interparental conflict in the neurotic group.

One possible interpretation of the results is that parental conflict and cross-sex parental dominance interact to produce a serious problem for the child. Several studies have shown that normal children tend to imitate the dominant parent regardless of sex (Bandura, Ross, & Ross, 1963; Hetherington, 1965). In a recent study using the RDT, Hetherington and Frankie (1967) found that parental conflict tended to increase imitation of the dominant parent. A strong identification with a parent of the opposite sex may lead to a variety of social and school adjustment difficulties (Kagan, 1964).

Nevertheless, the results relating the areas of parental conflict and the problems of the child suggest a more subtle process than simply the development of inappropriate sex-role behavior. The actual problem behavior of the child appears to be determined by the

nature of the conflict between the parents. Let us assume, along with Vogel and Bell, that the child is drawn into the husband-wife conflict. Now, if the child is the same sex as the dominant parent, he very likely follows the lead of that parent in resolving the conflict with no consequent sex-role problem, although with some alienation from the non-dominant parent. Depending on the characteristics of the dominant parent, such a child may show nonconforming behavior but is not likely to show up in a neurotic child guidance clinic population. On the other hand, if the child is of the opposite sex of the dominant parent he cannot follow the lead of that parent without entering into a sex-role conflict. Thus, a boy with a proachievement dominant mother may try to do well in school but fail because he does not want to be a sissy. At the same time he does not become a truant because this involves identification with the antiachievement nondominant father who is disparaged by the mother. While this explanation is speculative, it does account for the three major findings of the study—the cross-sex dominance pattern, the presence of hostile conflict, and the relation between parental conflict and child's problem.

REFERENCES

- ACKERMAN, N. W. *The psychodynamics of family life*. New York: Basic Books, 1958.
- BANDURA, A., ROSS, D., & ROSS, S. A. A comparative test of the status envy, social power, and secondary reinforcement theories of identification learning. *Journal of Abnormal and Social Psychology*, 1963, 67, 527-534.
- BECKER, W. C., PETERSON, D. R., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Factors in parental behavior and personality as related to problem behavior in children. *Journal of Consulting Psychology*, 1959, 23, 107-118.
- BOSZORMENYI-NAGY, I., & FRAMO, J. L. (Eds.) *Intensive family therapy*. New York: Harper-Hoeber, 1965.
- CLYDE, D., CRAMER, E. M., & SHERIN, R. J. *Multivariate statistical analysis*. Coral Gables, Fla.: Biometric Laboratory, University of Miami, 1966.
- FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology*, 1960, 61, 31-38.
- FARINA, A., & DUNHAM, R. M. Measurement of family relationship and their effects. *Archives of General Psychiatry*, 1963, 9, 64-73.
- FISHER, S., BOYD, I., WALKER, D., & SHEEN, D. Parents of schizophrenics, neurotics, and normals. *Archives of General Psychiatry*, 1959, 1, 149-166.
- FONTANA, A. E. Familial etiology of schizophrenia: Is a scientific methodology possible? *Psychological Bulletin*, 1966, 66, 214-227.
- FRANK, G. H. The role of the family in the development of psychopathology. *Psychological Bulletin*, 1965, 64, 191-205.
- HANDEL, G. Psychological study of whole families. *Psychological Bulletin*, 1965, 63, 19-41.
- HETHERINGTON, E. M. A developmental study of the effects of sex of the dominant parent on sex-role preference, identification, and imitation in children. *Journal of Personality and Social Psychology*, 1965, 2, 188-194.
- HETHERINGTON, E. M., & FRANKIE, G. Effects of parental dominance, warmth, and conflict on imitation in children. *Journal of Personality and Social Psychology*, 1967, 6, 119-125.
- KAGAN, J. Acquisition and significance of sex typing and sex role identity. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of Child Development Research*. Vol. 1. New York: Russell Sage Foundation, 1964.
- LATSCHA, R. Tests of significance in a 2×2 contingency table. *Biometrika*, 1953, 40, 74-86.
- LEVY, D. M. *Maternal overprotection*. New York: Columbia University, 1945.
- LIDZ, T., FLECK, S., & CORNELISON, A. R. *Schizophrenia and the family*. New York: International Universities Press, 1965.
- PETERSON, D. R., BECKER, W. C., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Parental attitudes and child adjustment. *Child Development*, 1959, 30, 119-130.
- SATIR, V. *Conjoint family therapy*. Palo Alto: Science and Behavior, 1964.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SPIEGEL, J. P., & BELL, N. W. The family of the psychiatric patient. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959.
- STRODTBECK, F. Husband-wife interaction over revealed differences. *American Sociological Review*, 1951, 16, 468-473.
- VOGEL, E. F., & BELL, N. W. The emotionally disturbed child as the family scapegoat. In N. W. Bell & E. Vogel (Eds.), *A modern introduction to the family*. Glencoe, Ill.: Free Press, 1960.
- WARNER, W. L. *Social class in America*. New York: Harper, 1960.
- WINDER, C. L. Some psychological studies of schizophrenia. In D. D. Jackson (Ed.), *The etiology of schizophrenia*. New York: Basic Books, 1960.
- ZUCKERMAN, M., BARRETT, B. H., & BRAGIEL, R. M. The parental attitudes of parents of child guidance cases: I. Comparisons with normals, investigations of socioeconomic and family constellation factors, and relations to parents' reactions to the clinics. *Child Development*, 1960, 31, 401-417.

(Received July 12, 1967)

SYMPTOM TREATMENT AND SYMPTOM SUBSTITUTION IN ENURESIS¹

BRUCE L. BAKER²

Yale University

This study explored the hypotheses that (a) Aspects of the therapist-patient relationship are responsible for successful behavior therapy; and (b) substitute problems will arise following remission of symptoms. Thirty enuretic children were treated either by a conditioning method or by methods devised to duplicate its motivational aspects. Adjustment measures on the enuretic Ss and control children were obtained independent of treatment. Conditioning was superior to other methods, suggesting that successful treatment was not based solely on the therapist-patient relationship. Subsequent to treatment, test measures did not indicate a decline in adjustment; on several measures, significant improvement was found.

Two questions are frequently posed to the behavior therapist. First, what are the total effects of treatment on the individual? Traditional psychiatry (e.g., Freud, 1959) views maladjustive behavior as a sign that there is an underlying disorder, and warns that even if a problem can be removed by a nondynamic therapy a relapse or another new problem will follow, since the basic causes have not been treated. This "symptom substitution" hypothesis is probably the principal theoretical reason why many clinicians hesitate to use behavior therapy. Yet while the vast majority of behavior-therapy follow-up reports indicate that no new problems have arisen following successful treatment (e.g., Baker, Kahn, & Weiss, 1968; Lang & Lazovik, 1963; Lazarus, 1961, 1963; Wolpe, 1961), there has been no well-controlled study of this hypothesis.

Second, why does behavior therapy work when it does? Although behavior therapists stress the role of the learning paradigm in treatment, the learning theory upon which

some of these methods are based has been criticized as outmoded (Breger & McGaugh, 1965). It seems possible that behavior therapy's success might be attributed to other aspects of the treatment, such as the therapist's enthusiasm, the kind of therapist-patient relationship which behavior-therapy methods establish, and the demands implicit in a "scientific" treatment.

The present study was concerned with both of the above questions. Behavior-therapy methods were employed to treat enuresis in children, and special attention was directed toward the efficiency of conditioning relative to other methods, and the changes in adjustment immediately following cure.

Bed-wetting seems a particularly appropriate disorder for research into the questions posed. Improvement can be accurately assessed, and the point of complete symptomatic relief is attainable and well defined. Also, traditional psychiatric conceptions of enuresis as a symptom of emotional disturbance (summarized in Mowrer, 1950) are quite different from the behavior therapist's view of enuresis as an isolated habit deficiency in the great majority of cases (e.g., Lovibond, 1964). Although the conditioning treatment of enuresis has been reported to be quite effective by many investigators (summarized in Lovibond, 1964), it is still not widely used, in part because of the symptom-substitution concern. For example, Spiering (1965) has recently written: "The removal of the symptom of enuresis, without providing other

¹ Based on a dissertation submitted to the faculty of the Graduate School of Yale University in partial fulfillment of the requirements for the PhD degree. The author wishes to thank Michael Kahn, his major advisor, and Fred D. Sheffield and Sidney J. Blatt, who served on his Advisory Committee. The author also wishes to thank Gilda Hymer for obtaining the schoolroom measures, and Irving H. Frank, Phillip Morse, and Ragaa Mazen for judging drawings. The conditioning devices were generously provided by Sears, Roebuck and Co.

² Now at the Department of Social Relations, Harvard University.

outlets for the child, leads to a replacement by other symptoms . . . [p. 30]." The few studies which have investigated this question of adjustment following cure (Baller & Schalock, 1956; Behrle, Elkin, & Laybourne, 1956; Biering & Jespersen, 1959; Lovibond, 1964) are inconclusive, however, because they were poorly (or not at all) controlled, for the most part lacked quantified measures, and did not assess change independent of the therapy.

The present study separated the treatment of enuresis and measurements of adjustment, so as to avoid the subtle forms of experimenter bias (Rosenthal, 1966) and demand characteristics (Orne, 1962) which continue to haunt practically all therapy-outcome studies. Measures of adjustment were taken by another *E* in the enuretic child's school, and these measures were taken under the guise of an entirely different research project.

METHOD

Subjects

The *Ss* were 90 elementary school children—30 enuretics and 60 controls. Two control *Ss* of the same sex as the enuretic *S* were selected randomly from each enuretic *S*'s classroom.

The enuretic *Ss* were obtained primarily in response to a newspaper article describing the project and inviting participation. A preliminary phone interview obtained the child's name, address, age, school and grade, and information about the enuresis. Also, the parent's report that in the family physician's opinion there was no organic problem was a prerequisite for being included. The enuretic sample consisted of 10 girls and 20 boys, with a median age of 8 yr. and a range from 6 to 12 yr. All but 4 of the children had been wetting since birth, and more than half were wet 7 nights a week.

Apparatus

The conditioning units used two foil pads, with holes in the top pad, separated by an absorbent sheet, and placed under *S*'s lower bed sheet. The pads were connected to a white plastic box which contained two 6-v. batteries, a sensitive relay, and a buzzer. Within seconds after the child began to wet, a circuit was completed and the buzzer sounded. The buzzer continued to sound until *S* got out of bed and shut it off.

The "gadgetry" device, used in the subsequent treatment of four waiting-list *Ss*, consisted of a bulletin board for a star chart, a container for stars, and a wind-up alarm clock, all mounted on a white wooden stand.

Procedure

The 30 enuretic *Ss* were arranged in triads according to the data obtained in the phone interviews, and 1 *S* was randomly assigned to each of three experimental conditions. Ten children were placed immediately in behavior therapy with a conditioning device (Group C). Ten children were concurrently given a wake-up treatment designed to duplicate all features of the conditioning treatment except the conditioning procedure (Group WU). Ten children were placed on a waiting list to begin treatment in the near future (Group WL). The Group WL families were scheduled for a brief home interview, in which *E* obtained further information about the child's enuresis.

Treatment

All treatment was carried out in the child's home. In the first visit, *E* told *S*'s parents that he was investigating both the process of cure and the changes which follow a child's becoming dry. Parents were asked to be particularly attentive to any changes in their child—especially new problems—both during and after treatment.

Group C. The *E* explained the operation of the conditioning device and encouraged the child to practice setting it up himself. The *S* was kept on the device until he had 14 consecutive days dry; the device was then taken off the bed but left in the house for an additional 14 days, after which *S* was designated as dry. If a child was not dry (or much improved) after 50 reinforcements (buzzer sounds), he was designated a failure. If a child later relapsed, he was begun again on the device when possible.

Group WU. The wake-up treatment was similar to the routine prescribed by many pediatricians, but with emphasis on regularity and thorough awakening (adapted from Smith, 1948). A fixed time was chosen before *S* usually wet, and the parents were told to awaken *S* every night at this time; the best time evolved during treatment. It was stressed that the child must be wide awake before going to the bathroom. When *S* had been dry for a week, he was not awakened for several nights during the next week. If he was dry on these nights, he was put on a new schedule whereby he would only be awakened the 2 nights following a wet night.

Otherwise, treatment was the same for conditioning and wake-up children. All parents aided the child in getting up at night and kept daily records of progress; all children kept progress records with star charts. Every child was visited at home once a week and phoned once a week by the author to review progress and offer suggestions. Parents in both groups were generally very cooperative in keeping records and appeared to be adequately following the treatment routine.

Group WL. Approximately 10 wk. after the original interview, *E* phoned the 10 waiting-list families. One *S* in Group WL had become dry and another *S* had not been examined medically after repeated requests; consequently, these 2 *Ss* were not

begun in treatment. To explore some peripheral hypotheses, the remaining 8 waiting-list Ss were now treated in the following way.

It seemed possible that the conditioning method might be effective because *E* brings a "gadget" into the home, thus in some way increasing the child's motivation. To obtain pilot data on this question, four Ss were placed in wake-up treatment, but were also brought the gadget described above, with instructions to place it beside the bed and to set the alarm clock at the time designated for awakening. The other four Ss received the same conditioning treatment as the original Group C.

Also, as an exploratory means of assessing the importance of the therapist's presence, there were two levels of contact with *E*, full and partial. Two Ss in each group received the same amount of contact with *E* as Group C and Group WU Ss had received (full contact: one visit and one phone call each week). The remaining Ss received only half as much contact with *E* (partial contact: a visit one week, a call the next week, etc.).

Measures of Adjustment

Parent measures. Parents were periodically asked whether they had observed any changes in their child, and they completed two rating scales, both before treatment began and several weeks after treatment was terminated. The Adjective Check List (ACL) was adapted from a scale used by Sarason, Davidson, Lighthall, Waite, and Ruebush (1960) and included 20 items designed to assess personality attributes such as confidence, anxiety, and responsibility. The scale included the 12 Sarason items which could readily be scored for adjustment and 8 new items. Items were presented in the form (Sad 321 123 Cheerful), and parents were asked to select the word which better typified *S* and to indicate the extent (a little, definitely, very much) by circling one number. The Behavior Problem Record (BPR) was a checklist of 26 specific childhood problems; the first 15 items were a scale devised by Lovibond (1964) in his work with enuresis, and the last 11 were added by the author. Many of the problems in the BPR have been associated with enuresis in the psychiatric literature or have been suggested as possible new symptoms if enuresis is removed without treating an underlying cause (e.g., fire-setting, thumbsucking, temper outbursts).

Teacher ratings. The enuretic Ss in the study were distributed throughout 23 public schools and 2 parochial schools; since no two enuretic Ss happened to be in the same classroom, 30 teachers were involved in the ratings. From each enuretic S's class list, two other children of the same sex were randomly selected as controls.

Although *E* initially contacted the superintendents and principals, all contacts with the teachers and all testing of the children were performed by another *E* (*E*₂). The research was presented to the teachers as a study of creativity. No mention ever was made of bed-wetting. Teachers were told that their class and these children had been selected at random. The *E*₂ explained that many of the hypotheses in the project

would be biased if the teachers were made aware of them, but that after the data were all gathered she would welcome the chance to meet with the teachers and explain more fully the project and how their work fit into it. Thus, only when the study was completed did teachers learn of the enuresis treatment and their relation to it.

The teachers were given a 67-item rating form for each of the three children. Each item was to be rated on a scale of either 1-5 or 1-7. Most of the items were taken from the Devereaux Elementary School Rating Scale (Spivack & Swift, 1966) and the Devereux Child Behavior Rating Scales (Spivack & Levine, 1964). Both scales had been factor analyzed, and in devising the teachers' rating form items were drawn from each of the factors. These factors represented either aspects of school behavior (e.g., academic anxiety, dependence on the teacher for learning) or personal maladjustment (e.g., social aggression, fears, inability to delay). Since most of the items were stated negatively and measured "problems," new items were added exploring such things as creativity, happiness, confidence, responsibility, maturity, and capacity for play. Items were stated in terms of overt behavior, and teachers were asked to compare *S*'s current behavior with that of other children his age.

Teachers were asked to complete these questionnaires three times throughout the year. Test I was just before treatment began for the enuretic *S*, Test II was about 10 wk. later, and Test III about 12 wk. after Test II.

Child measures. At these same times, the enuretic child and one of the two control children in each class were tested by *E*₂, who was not aware of the specific hypotheses being investigated and did not know which child was enuretic. It was announced to the class that *E*₂ wanted to find out some things about elementary school children and that two names had been picked "out of a hat." The test battery began with the Draw-A-Person and Draw-Your-Family Tests and included four TAT cards (not to be reported here) and a self-report questionnaire in two parts: (a) the Self-Image Questionnaire, a 16-item scale designed by the author to measure feelings which might arise from being enuretic and including positively and negatively stated items such as, "My parents are very happy with me," and "I often do things I wish I had never done"; and (b) the Neurotic Inventory, a 20-item scale measuring neurotic problems devised and employed with enuretics by Lovibond (1964) and including items such as, "I worry quite a bit about things which might happen," and "Most of the time I feel down in the dumps." Items in both scales were presented verbally to *S* as "things which boys [girls] I know have said," and *S* was asked if he was like that child. (E.g., "Bob said: I get a lot of headaches. Are you exactly like Bob? Somewhat? Not at all?")^a

^aAfter the first testing, it seemed that some Ss might be responding defensively to the questionnaire scales, and the need for a defensiveness measure was realized. Hence, the D scale employed by Sarason et

Neither the parents of tested Ss nor Ss themselves had any knowledge of the relationship between this school testing and treatment; if parents inquired at school about the testing, they too were told that their child was in a creativity study of randomly selected children.

After the final testing it was necessary to determine which of the control Ss happened to be enuretic. The *E* phoned the parents of the control Ss requesting that they complete a questionnaire about their child. Parents were sent the same questionnaire which the parents of enuretic Ss had filled out, except that two additional items were added to the BPR: bed-wetting and daytime wetting.

RESULTS

Reliability and Intercorrelation of the Scales

Table 1 shows the Spearman-Brown corrected split-half reliabilities of the questionnaire measures for Test I and the test-retest coefficients. These reliabilities are satisfactory for questionnaires of this type. Drawing variables requiring subjective judgment were scored independently by the author and another judge, with Test I interjudge reliabilities ranging from .43 to .95. For analyses of drawing variables, an average of the two judges' ratings was used.

Comparison of Enuretics and Controls on Pretreatment Measures

Enuretics and controls did not differ significantly on any item in the self-report scales, any factor in the teachers' ratings, or either total score. None of the specific drawing variables in either the Draw-A-Person or Draw-Your-Family Tests differentiated enuretics from controls. Also, two clinical psychologists making blind judgments of enuretic-control pairs of drawings were unable to identify the enuretics' drawings any better than chance expectation.

Treatment

Figure 1 summarizes the results of the first 10 wk. of treatment. There was a sizable initial drop in wetting frequency from the prelevel to the first 2 wk. of treatment for both treatment groups. Following this, the conditioning group frequency decreased over the 10-wk. period, while the frequency of wetting in wake-up Ss remained essentially unchanged.

al. (1960) was administered to most Ss in Test II and to all Ss in Test III.

TABLE 1
SPEARMAN-BROWN CORRECTED SPLIT-HALF RELIABILITIES OF THE QUESTIONNAIRES FOR TEST I AND TEST-RETEST RELIABILITIES

Scale	Corrected r	Tests I-II ^a	Tests I-III
Children's Neurotic Inventory	.94	.80	.70
Children's Self-Image Questionnaire			
Negative items ^b	.79	.74	.67
Teachers' rating form	.85	.86	.83
Adjective Check List			
Mother	.89	.87	—
Father	.83	.92	—
Behavior Problem Record			
Mother	.89	.62	—
Father	.93	.71	—

^a For the child and teacher measures, there was an average of 10 wk. between Tests I and II and an average of 22 wk. between Tests I and III. For the parent measures, the questionnaire completed just before treatment was correlated with the measure following treatment, yielding only one coefficient.

^b The positive self-image items were unreliable and did not correlate with the negative items or other measures; consequently, only the negative items were considered.

This Conditions \times Weeks interaction was significant at $p < .05$ ($F = 2.54$, $df = 4/72$) by a Lindquist (1956) Type I analysis of variance of the treatment data (excluding the prelevel). Comparing the number of wet nights in Weeks 1-2 and Weeks 9-10, a second-order t test found the improvement in Group C to have been significantly greater than in Group WU ($t = 2.89$, $df = 18$, $p < .01$). These results indicate that the two conditions produced differential effects over time. By Weeks 9-10, the conditioning group

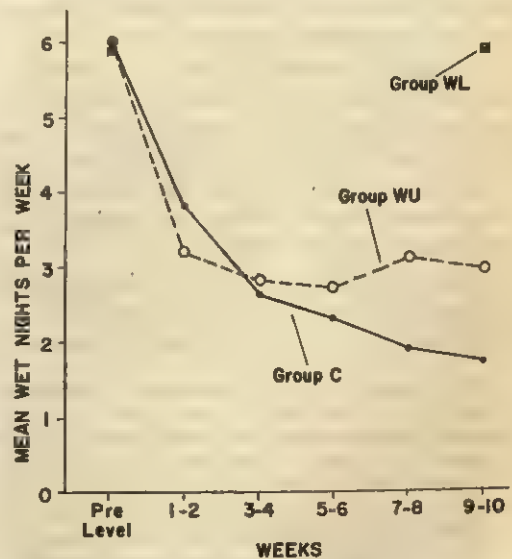


FIG. 1. Mean number of wet nights a week over the first 10 wk. of treatment for Groups C, WU, and WL.

was wetting significantly less than the wake-up group ($t = 1.70$, $df = 18$, $p = .05$).⁴

An interview with waiting-list families an average of 10.6 wk. after *E*'s first visit found the reported mean frequency of wet nights unchanged. During this period, one Group WL S had become dry, but his previous frequency had been only 2 wet nights a week, the lowest in the study, and his decrease was offset by two Ss whose frequency increased. Contrasting the number of wet nights in the previous 3 wk. with the comparable 3 wk. for Group C and Group WU (Weeks 8–10 of treatment), the mean number of wet nights for Group WL ($\bar{X} = 17.7$) was significantly greater than either the mean for Group C ($\bar{X} = 5.5$; $t = 4.30$, $df = 18$, $p < .001$) or the mean for Group WU ($\bar{X} = 9.3$; $t = 3.21$, $df = 18$, $p < .01$).

The improvement during the first 8 wk. of subsequent conditioning or "gadgetry" treatment for the waiting-list children was analyzed by an analysis of variance. While the conditioning treatment was more effective than the gadgetry control, this difference was not statistically significant. (The Ss with the gadget did no better than previous wake-up Ss, but Ss on conditioning did not do as well as previous conditioning Ss.) Children who had been seen less frequently (partial contact) showed more improvement than Ss who received full contact, though this difference was not statistically significant.

Treatment summary. In all, 14 Ss were begun on the conditioning device; of these, 11 attained initial arrest. Fourteen Ss were begun in wake-up conditions; of these, 2 attained initial arrest. Ten Ss who had shown little improvement were switched to conditioning, and 7 subsequently became dry.

Of the 30 enuretic Ss, then, 1 became dry without treatment, 1 was not begun in treatment, and 1 was terminated before treatment

⁴ In evaluating treatment differences, it is important to note that Group WU Ss had the aid of being awakened during the night until they had progressed sufficiently to be left on their own, whereas Group C Ss did not. The implication is that Group WU Ss appear to be more improved than was actually the case. For example, six Group WU Ss had improved sufficiently to be left on their own some nights. The mean percentage of wetting was 22.5 on nights when they were awakened and 38.5 on nights when they were not awakened ($t = 2.51$, $df = 5$, $p < .03$).

could be completed. Of the remaining 27 Ss, 74% were "cured" (dry for at least a period of 1 mo.), and an additional 15% were "very much improved" (dry for at least a 2-wk. period and wetting less than 1 night a week). During the follow up, which averaged 6 mo., 4 cured Ss relapsed, although 2 of these who were retreated became dry again.

Changes Following Treatment

Parents' interviews and questionnaires. It is important to note in assessing the results of parents' measures that changes similar to those reported for enuretic Ss might also have been reported for control Ss if such comparable measures were available. In addition, parents' measures may reflect a desire to please *E*. It was for these reasons, of course, that independent school measures were taken on both enuretic and control Ss.

The most frequently reported observation of parents was the child's happiness at becoming dry. Many children were able for the first time to sleep overnight with friends or relatives or to go to summer camp; three boys immediately joined the Boy Scouts. Children were reported to be venturing into new activities, taking more responsibility, and becoming more autonomous. For instance, parents made such statements as, "She seems able to do things for herself"; "He started a paper route"; "Now she's cooking, sewing, reading."

While most parents reported that they had observed no new problems, several parents did report a new problem at some time. One boy defecated in his pants several times immediately following treatment. Another began to depend more on his mother to make decisions for him. A third boy developed an "eyeblick." None of these proved to be lasting, and each seemed to arise from specific new stresses, although it is not possible to say if they had any relation to the treatment.

Both mothers and fathers of cured Ss indicated improvement in adjustment on the ACL and the BPR. Combined parents' scores on the ACL showed a mean improvement of 3.5, significant at $p = .002$. Combined parents' scores on the BPR showed a mean improvement (decrease in reported problems) of -2.8 , significant at $p = .05$.

School measures. Test II afforded a measure

of changes which occurred either during treatment or immediately thereafter. There was an average of 10 wk. from Test I to Test II, at which time 17 Ss were judged improved (wetting less than half their original frequency). There was an average of 22 wk. from Test I to Test III. By Test III, 13 Ss were cured, and another 8 Ss were very much improved (wetting once a week or less). Children designated as cured had not been wetting for an average of almost 3 mo.

Teacher ratings. Change scores were determined for total teachers' ratings from Test I to Test II; enuretics were divided into "improved" and "unimproved" and compared with each other and with their controls. For Test I to Test III comparisons, enuretics were classified as cured, very much improved, or slightly improved and compared with their respective controls. None of the mean change scores, nor the differences between them, approached statistical significance; teachers showed a high degree of consistency in their ratings, and mean change scores were small. Change scores on only 3 of the 24 factors significantly differentiated cured enuretics from their controls, with cured enuretics increasing significantly more ($p < .05$) in academic anxiety, drive for academic success, and unethical behavior.⁵

Children's self-report questionnaires. There was a general downward trend (fewer reported problems) on the questionnaires from Test I to Test II. The direction of changes for enuretics and controls is shown in Table 2. While treatment was in progress or only recently terminated, enuretic Ss who had shown improvement in bed-wetting reported fewer problems than previously. This improvement compared favorably with changes shown in questionnaire scores for both unimproved enuretic Ss and control Ss.

In analyzing changes from Test I to Test III, special attention was given to Ss designated as cured,⁶ as these Ss would be most

TABLE 2
DIRECTION OF CHANGE IN QUESTIONNAIRE SCORES
FROM TEST I TO TEST II

Questionnaire	Improved enuretics	Unimproved enuretics	Controls
Neurotic Inventory			
Increase	0	3	10
Same	0	3	1
Decrease	17	7	19
Self-Image Questionnaire ^b			
Increase	2	4	6
Same	1	1	1
Decrease	14	8	23

Note.—Decrease indicates improvement (fewer reported problems).

^a $\chi^2 = 15.19, p < .01$.

^b $\chi^2 = 2.14, p > .10$.

likely to evidence adverse changes according to the symptom-substitution hypothesis. On the Neurotic Inventory, no cured S reported more problems, and on the Self-Image Questionnaire only one S increased in score. The mean improvement for cured enuretics was greater than the mean improvement for their controls on both the Neurotic Inventory ($t = 1.49, df = 20, p = .08$) and the Self-Image Questionnaire ($t = 1.79, df = 20, p < .05$). Combining the questionnaires, the mean improvement from Test I to Test III for cured enuretics ($\bar{X} = -15.6$) was significantly greater than for their controls ($\bar{X} = -6.4$) at $p = .03$ ($t = 1.93, df = 20$).⁷

There were no questionnaire items on which cured and very much improved enuretics showed a significant change for the worse or their controls improved more than the enuretic Ss. Cured and very much improved enuretics improved significantly more than did the controls ($p < .05$) on the following items: "I'm always being scolded or punished by someone"; "I've often been punished for nothing"; "I often feel sick in the stomach"; "Most of the time I feel down in the dumps."

Drawings. Scoring of specific variables on the Draw-A-Person and Draw-Your-Family

school. The control of another cured S had moved away, so this pair was not included. In all, then, there were 11 cured-control pairs.

⁷ One explanation for this decrease would be increased defensiveness. Since no D-scale scores were available for Test I, the only change which could be analyzed was from Test II to Test III. The defensiveness scores were fairly stable, and the changes did not differentiate cured enuretics and controls.

⁵ Actually, cured enuretics only increased somewhat in drive for academic success and unethical behavior, but controls decreased in score on these variables, resulting in the significant difference. However, there was a large Test I difference for these Ss on these factors, and even after the above changes the control Ss still scored higher than enuretics.

⁶ One cured S and his control could not be tested a third time due to lack of cooperation from the

Tests failed to reveal significant changes. A total score was derived for each *S* on the former, based on conceptions in the drawing literature (e.g., Machover, 1951; Mazen, 1963) of what constitutes an improvement in adjustment. To weight each variable equally, only the sign of change was considered. Signs of a shift toward better adjustment were an increase in number of body parts, size of the person, number of colors used, appropriateness of colors, amount of clothing, smile and movement, and a decrease in erasures, pressure, distortion, bizarreness, paper chopping, transparency, stick figures, drawing side views, drawing the opposite sex, and negative affect. Cured enuretics showed a mean improvement ($\bar{X} = 1.36$, $p = .06$) and differed significantly from their controls who showed a slight worsening ($\bar{X} = -1.45$; t , second-order difference = 2.29, $df = 20$, $p = .02$).

Also, two clinical psychologists with the diplomate and 15 and 30 yr. of experience in therapy and diagnosis were asked to evaluate the Draw-A-Person Test drawings for expressions of maladjustment. Each of the 30 enuretic and 30 control *Ss* was randomly assigned a number between 1 and 60, and each *S's* three drawings (Tests I, II, and III) were randomly ordered in one of the six possible permutations. Judges were told the age, race, and sex of each child and were to rank order the three drawings with respect to overall adjustment. In assessing the drawings, then, a judge knew neither which drawings had been done by enuretic and nonenuretic *Ss* nor the order in which they had been drawn.

In this type of analysis, differences between groups are reflected in the slope of the curves connecting mean ranks across testings. As seen in Figure 2, for each judge the mean ranks for the cured enuretics decrease across the three tests, whereas the control slope is essentially flat. The improvement in adjustment rank from Test I to Test III for cured enuretics was significant for both Judge 1 ($t = 3.81$, $df = 10$, $p = .002$) and Judge 2 ($t = 2.36$, $df = 10$, $p = .02$). When this Test I-Test III change score for cured enuretics was compared with a similar score for their controls, the second-order t test showed that improvement in adjustment for cured enuretics was significantly greater than for con-

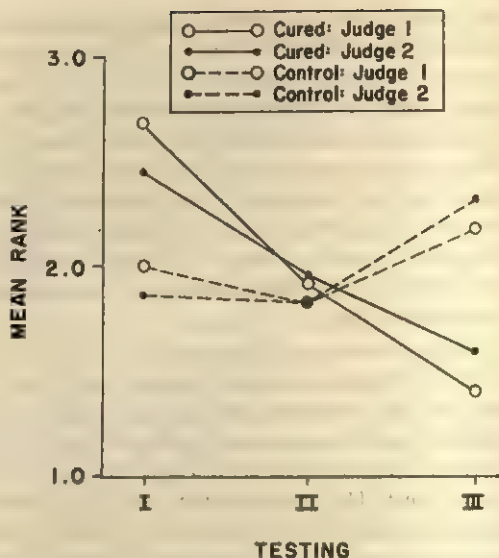


FIG. 2. Judges' ranks of overall adjustment in Draw-A-Person Test drawings; mean ranks for each judge are plotted over the three testings for cured enuretic *Ss* and their control *Ss*. (Lower ranks indicate better adjustment.)

trols—for Judge 1, $p = .007$ ($t = 2.72$, $df = 20$) and for Judge 2, $p = .004$ ($t = 2.88$, $df = 20$).

DISCUSSION

The present study suggests, then, that it is extremely doubtful that the successful conditioning treatment of bed-wetting can be explained as some kind of "transference cure," based on only the therapist-patient relationship and the motivation which it engenders. The wake-up procedure was not as effective as the total behavior-therapy treatment, despite an equal amount of therapist contact. These findings are similar to those of Lang and Lazovik (1963), Lazarus (1961), Paul (1966), and Davison (1968) who contrasted systematic desensitization to other methods with equal therapist contact.

Another approach to this question of the relative importance of the therapist and his technique is to withdraw the therapist to some extent and to see if treatment becomes less effective. The exploratory findings indicated that treatment is just as effective with less therapist contact, a result consonant with the fact that many parents buy conditioning de-

vices and cure their child's enuresis with no professional help at all.

The measures of adjustment did not show a worsening in adjustment subsequent to treatment; rather, other improvements were found. Children were reported to be happier, less anxious, and more grown-up, assuming responsibility and venturing into new activities. The changes on childrens' self-report items about punishment suggest that the cure may have had beneficial effects on the parent-child relationship or at least the child's perception of it. Finally, the drawings showed a dramatic improvement in what might be termed "self-image."

Whereas enuresis is considered to be a symptom of emotional disturbance by many clinicians, traditional predictions about treatment and subsequent adjustment for such symptoms were not upheld. This might suggest that the symptom-substitution hypothesis is untenable and that neurotic symptoms, in the traditional dynamic sense, can be removed with equanimity despite the psychological gains they afford. Perhaps in those cases where treatment succeeds, the benefits derived from being dry outweigh such "primary gains." On the other hand, it is possible that new symptoms did not arise because bed-wetting is simply a habit deficiency rather than an expression of, and outlet for, internal conflict. In any case, the dangers of a direct treatment of enuresis seem to have been overstated, and similar research on other classical disorders might be of considerable value in further understanding the symptom-substitution issue and the more basic question of symptom formation.

REFERENCES

- BALLER, W., & SCHALOCK, H. Conditioned response treatment of enuresis. *Exceptional Child*, 1956, 22, 233-236.
- BEERLE, F. C., ELKIN, M. T., & LAYBOURNE, P. C. Evaluation of a conditioning device in the treatment of nocturnal enuresis. *Pediatrics*, 1956, 17, 849-855.
- BIERING, A., & JESPERSEN, I. The treatment of enuresis nocturna with conditioning devices. *Acta Paediatrica*, 1959, 48(Monogr. Suppl. No. 118), 152-153.
- BREGER, L., & MCGAUGH, J. L. Critique and reformulation of "learning-theory" approaches to psychotherapy and neurosis. *Psychological Bulletin*, 1965, 63, 338-358.
- DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, 73, 91-99.
- FREUD, S. On psychotherapy. (Orig. publ. 1904) In, *Collected papers of Sigmund Freud*. Vol. 1. New York: Basic Books, 1959.
- KAHN, M., BAKER, B. L., & WEISS, J. M. Treatment of insomnia by relaxation. *Journal of Abnormal Psychology*, 1968, 73, 556-558.
- LANG, P. J., & LAZOVIC, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology*, 1963, 66, 519-525.
- LAZARUS, A. A. Group therapy of phobic disorders by systematic desensitization. *Journal of Abnormal and Social Psychology*, 1961, 63, 504-510.
- LAZARUS, A. A. The results of behaviour therapy in 126 cases of severe neurosis. *Behaviour Research and Therapy*, 1963, 1, 69-80.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
- LOVIBOND, S. H. *Conditioning and enuresis*. Oxford: Pergamon Press, 1964.
- MACHOVER, K. Drawings of the human figure. In H. Anderson & G. Anderson (Eds.), *An introduction to projective techniques*. Englewood Cliffs, N. J.: Prentice-Hall, 1951.
- MAZEN, R. Comparative study of the drawings of normal and disturbed children. Unpublished manuscript, Yale University, 1963.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
- ORNE, M. T. On the social psychology of the psychological experiment with particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- PAUL, G. L. *Insight versus desensitization in psychotherapy*. Stanford: Stanford University Press, 1966.
- ROSENTHAL, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.
- SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., WAITE, R. R., & RUEBUSH, B. K. *Anxiety in elementary school children*. New York: Wiley, 1960.
- SMITH, S. *The psychological origin and treatment of enuresis*. Seattle: University of Washington Press, 1948.
- SPEERLING, M. Dynamic considerations and treatment of enuresis. *Journal of the American Academy of Child Psychiatry*, 1965, 4, 19-31.
- SPIVACK, G., & LEVINE, M. The Devereux Child Behavior Rating Scales: A study of symptom behaviors in latency age atypical children. *American Journal of Mental Deficiency*, 1964, 68, 700-717.
- SPIVACK, G., & SWIFT, M. The Devereux Elementary School Rating Scale: A study of the nature and origin of achievement related to disturbed classroom behavior. *Journal of Special Education*, 1966, 1, 71-90.
- WOLFE, J. The systematic desensitization treatment of neuroses. *Journal of Nervous and Mental Disease*, 1961, 132, 189-203.

(Received July 14, 1967)

ORAL IMAGERY AND ALCOHOLISM

SHARON BERTRAND AND JOSEPH MASLING

State University of New York at Buffalo

Twenty Veterans Administration patients with histories of heavy daily alcoholic consumption were matched on age, education, IQ, and diagnosis with 20 patients with no history of reliance on alcohol, and their scores on a Rorschach measure of orality were compared. As predicted, the alcoholics gave more oral dependent responses than the controls ($p = .01$). Although total scores on oral sadism did not discriminate between groups, the alcoholic Ss gave more responses on two of the oral sadistic subcategories, burdens and overwhelming figures. The striking similarity of subcategory performance of this group to an earlier group of obese Israelis was discussed.

At one time or another almost every theory of personality, ranging from psychoanalysis to learning theory, and almost every method of data collection, ranging from anthropological field studies to factor analysis of MMPI items, has been used to try to account for the etiology of alcoholism. This conglomeration has produced a mixture of clinical observations, unrelated data, and free-floating speculation. Perhaps the most frequently suggested predisposing factor is a state of unsatisfied oral dependent needs. Fenichel (1945), for example, described alcoholics as being "characterized by their oral and narcissistic premorbid personalities, usually as a result of difficult family constellations in which specific oral frustrations are created [p. 379]." A standard psychiatric text states that "evidence of residual oral traits which are often seen in adults include such things as alcoholism . . . [English & Finch, 1954, pp. 17-18]."

The importance of oral dependent needs in the etiology of alcoholism has received partial support from anthropological and experimental studies. Bacon, Barry, and Child (1965) in reviewing 110 preliterate societies concluded that heavy consumption of alcohol results in part from unsatisfied dependency needs. In the United States a longitudinal study of alcoholics noted that the effects of alcohol and the act of drinking help to resolve conflict between dependency needs and the pressure from society to be independent (McCord & McCord, 1960). The greater use of field-dependent solutions by alcoholics has been repeatedly demonstrated (Goldstein & Chotlos, 1965; Karp & Konstadt, 1965; Karp,

Poster, & Goodman, 1963; Karp, Witkin, & Goodenough, 1965; Witkin, Karp, & Goodenough, 1959). Wiener (1956) compared positive, neutral, and negative oral responses on the Rorschach test of alcoholics and neurotic depressives and found no differences between groups in the total number of oral percepts, although the alcoholics gave significantly more positive oral responses, and the depressives gave more negative oral responses.

The oral stage of development is typically seen as progressing through two periods—an oral dependent phase, characterized by sucking, dependence, passivity, and immaturity, and an oral sadistic phase, characterized by biting, chewing, and ambivalence (Blum, 1953). Psychoanalytic hypotheses about alcoholism have emphasized only the oral dependent phase, not the oral sadistic phase. Measures derived from the Rorschach test of the two phases of the oral need have demonstrated validity in discriminating between an obese and nonobese population (Masling, Rabie, & Blondheim, 1967) and between students who yielded and those who did not in an Asch-type conformity experiment (Masling, Weiss, & Rothschild, 1968).

The hypothesis of this study is that alcoholic Ss will give more oral dependent responses on the Rorschach test than a control group of nonalcoholic Ss. No differences are hypothesized for oral sadistic responses.

METHOD

Subjects

The Ss were 40 Caucasian male veterans receiving treatment in the psychiatric unit of a Veterans Ad-

TABLE 1
DESCRIPTORS OF EXPERIMENTAL AND
CONTROL SUBJECTS

Variable	Experi- mental group	Control group
Age		
Range	24-60	26-58
<i>M</i>	39.25	40.30
<i>σ</i>	8.20	8.06
IQ		
Range	87-135	90-127
<i>M</i>	106.75	107.15
<i>σ</i>	13.79	10.12
Education (no. yr.)		
Range	5-17	8-16
<i>M</i>	11.75	11.25
<i>σ</i>	3.51	2.19
Diagnosis		
Schizophrenic		
Total	11	11
Chronic undifferentiated	6	4
Paranoid	4	5
Schizoaffective	1	1
Catatonic	0	1
Personality disorders		
Total	4	4
Schizoid personality	1	2
Passive-aggressive		
Personality	3	2
Psychoneurotic		
Total	5	5
Anxiety reaction	1	2
Conversion reaction	0	1
Depressive reaction	4	2

ministration hospital.¹ The experimental group consisted of 20 patients whose therapists had classified them as being strongly dependent upon alcohol as a means of coping with their life situations; while none had a primary diagnosis as an alcoholic, all had a history of heavy daily alcoholic consumption throughout their adult lives. Their drinking was uncontrollable, continual, and interfered seriously with their vocational and social adjustment. The control Ss were selected from the same psychiatric unit and were either total abstainers from the use of alcohol or were light social drinkers. Control and alcoholic Ss were matched for age, number of years of education, scores on the WAIS, and diagnosis.

Testing Procedure

All Ss were given individual psychological tests within the first 3 wk. of hospitalization. Typically a patient who entered the hospital showing signs of prolonged drinking would be allowed several days to "dry out" before taking the tests. The 40 Ss were tested by 22 different examiners. Each S was given

¹ The authors are indebted to Martin Staiman, chief psychologist at the Buffalo Veterans Administration Hospital, for permission to use patient test records.

an individual Rorschach, but only the free-association responses were used in this study.

Scoring for Oral Imagery

The definitions of oral dependence and oral sadism were adapted from Schafer (1954). In this simple scoring method the subcategories of oral dependence and oral sadism are defined and examples given. The following subcategories were used: oral dependence—foods, food sources, food objects, food providers, passive food receiver, food organs, supplicants, nurturers, gifts and gift givers, good luck, oral activity, needing help or being passive, pregnancy and reproductive anatomy; oral sadism—destruction, aggression, or maiming, overwhelming figures, figures that deprive, deprivation, faulty oral capacity, oral assault, burdens.

To prevent contamination of the criterion by alcoholic responses on the Rorschach, a separate category was established for all alcohol, drinking, and liquid percepts.

RESULTS

Reliability

The Rorschach tests, with all identification removed, were scored independently by two judges.² There were 1,047 responses on the 40 Rorschach protocols; of this number there were 43 disagreements on the presence or absence of an oral percept, producing 96% agreement on this dimension. The disagreements were resolved by the two judges, again without knowledge of the identity of Ss. Once agreement was obtained on the presence of an oral percept, the judges agreed on 96% of the responses assigned to the oral dependent category and on 93% of the responses assigned to the oral sadistic category.

Total Number of Responses and Oral Percepts

The correlation between the number of responses each S produced and the number of oral responses he gave was .62 ($p = .005$); the correlation between the number of oral dependent responses and oral sadistic responses was .32 ($p > .05$).

Comparison of Alcoholic and Control Subjects

The mean number of Rorschach responses was 31.05 for the alcoholic Ss and 21.30 for the controls, a difference significant at the .02 level ($t = 2.52$). In order to control for the effect of this difference in productivity, the

² The raters were the first author and Lillie Weiss.

oral percepts for each *S* were divided by the total number of responses he gave, so that the basic score for each *S* was expressed as a proportion. Using proportions, and excluding alcohol and drinking responses,³ it was found that the alcoholic *Ss* gave more oral dependent responses than the controls ($\chi^2 = 6.40$, $p = .01$, one-tailed test). The two groups did not differ in the overall use of oral sadistic responses.

Of the 13 oral dependent subcategories, there were significant differences between groups on 2; compared with the responses of the control *Ss*, a greater percentage of alcoholic responses was in the food subcategory ($p = .05$) and food organ subcategory ($p = .05$). While there were no differences between groups on the overall oral sadistic category, the alcoholics gave a greater percentage of responses than the controls in 2 of the 7 subcategories—burdens ($p = .02$) and overwhelming figures ($p = .05$). Table 2 presents the distribution of responses for the alcoholic and control *Ss* on those subcategories in which significant differences were found. (For purposes of comparison, Table 2 also lists subcategory differences found in an earlier study of Israeli obese—Masling, Rabie, & Blondheim, 1967.)

Inspection of the individual test records uncovered a surprising relationship: the significant differences discussed above were due primarily to the response patterns of 10 of the alcoholics. Every one of these 10 *Ss* gave a burden response (e.g., a camel, two men trying to lift a heavy object) as well as an overwhelming figure response (e.g., an alligator with his mouth open, a witch); the pairing of these two responses was found neither in the remainder of the alcoholic *Ss* nor in the control *Ss*. In addition, this subsample of 10 alcoholic *Ss* gave more oral dependent responses ($M = 3.6$) than either the other alcoholic *Ss* ($M = 2.0$) or control *Ss* ($M = 1.0$). When the background factors of these 10 *Ss* were examined, it was found that they tended to be older, brighter, and better educated than the other *Ss*, although none of the differences was significant.

³ There were four such responses given, all by alcoholic *Ss*.

TABLE 2

COMPARISON OF ALCOHOLIC, OBESE, AND CONTROL SUBJECTS ON SELECTED SUBCATEGORIES OF RORSCHACH RESPONSES

Response subcategory	Alcoholic study: No. records with above median subcategory responses ^a		Obesity study: Distribution of subcategory responses ^b	
	Alcoholic <i>Ss</i> ^c	Control <i>Ss</i> ^c	Obese <i>Ss</i> ^c	Control <i>Ss</i> ^c
Oral dependence				
Food	9**	3	Not scored	Not scored
Food organs	8**	2	8****	0
Nurturers	1	0	7**	1
Supplicants	0	1	6*	1
Oral sadism				
Overwhelming figures	10**	4	14****	1
Burdens	10***	3	8*	2

^a Probabilities determined by chi-square.

^b Probabilities determined by Fisher exact test. Study by Masling, Rabie, and Blondheim (1967).

^c $N = 20$.

^d $N = 16$.

** $p < .10$.

*** $p < .05$.

**** $p < .02$.

***** $p < .01$.

DISCUSSION

These results support clinical observations of the oral characteristics of alcoholics and also partially confirm the earlier results of Weiner (1956) who found that alcoholics give more positive oral responses (essentially the oral dependent category of the present study) than depressive *Ss*. While the hypothesis was confirmed, this type of correlational investigation can obviously make no statement regarding cause and effect; from the evidence reported here it is as reasonable to conclude that oral dependence results from an alcoholic condition as it does to conclude that it produces it.

Two findings were especially impressive, because both were unexpected. Within the alcoholic group, the tendency to give oral responses varied considerably, with a subgroup of 10 reporting many more oral percepts than the others. Not only did these 10 report more oral responses, but they also paired two types of images—overwhelming figures and burdens—which on the surface seem to have few common elements. There are many kinds of alcoholic personality patterns (Jones, 1963; Pollner, 1965), and this study may have inadvertently isolated one

such group which represents oral dependent needs in a relatively strong manner. The pairing of overwhelming figures and burden responses in these 10 Ss suggests that these percepts may share some symbolic meaning for those with an oral pathology.

The second unexpected result was the close parallel in the subcategory patterns in this American male alcoholic population and in a predominantly female population of obese Israeli (Masling, Rabie, & Blondheim, 1967). In both groups the clinical populations gave significantly more food organ responses, overwhelming figure responses, and burden responses than their respective control groups. When groups of different sex, different religious and national origin, and different pathology show similar test patterns, it is likely that a meaningful phenomenon is involved. Logically this resemblance could arise if the obese were alcoholic or if the alcoholics were obese. Unfortunately, quantitative evidence is not available on this point, but experienced observers all report the almost total absence of alcoholism in either the Jewish or Moslem residents of Israel (Jones, 1963).⁴ Of course, the American alcoholics could be obese, but the staff members of this particular Veterans Administration hospital have reported that many Ss used in this study were so severely dependent on alcohol that they ate irregularly and poorly.

The similarity between alcoholic and obese Rorschach performance may be related to early deprivation of oral gratification in infancy and subsequent fixation at the oral level, as suggested by psychoanalysis. Perhaps it is related to the learning in infancy of oral behavior as a means of reducing anxiety, as a learning theory would suggest. In any event, the present study on alcoholics in conjunction with the earlier study on obesity has estab-

lished a parallel in responding to inkblots of two groups who have only one characteristic in common—an overuse of food or drink.

REFERENCES

- BACON, M. K., BARRY H., & CHILD, I. A cross-cultural study of drinking: II. Relations to other features of culture. *Quarterly Journal of Studies on Alcohol*, 1965, (Monogr. Suppl. 3), 29-48.
- BLUM, G. *Psychoanalytic theories of personality*. New York: McGraw-Hill, 1953.
- ENGLISH, O. S., & FINCH, S. M. *Introduction to psychiatry*. New York: Norton, 1954.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- GOLDSTEIN, G., & CHOTLOS, J. Dependency and brain damage in alcoholics. *Perceptual and Motor Skills*, 1965, 21, 135-150.
- JONES, H. *Alcoholic addiction, a psycho-social approach to abnormal drinking*. London: Tavistock Publication, 1963.
- KARP, S., & KONSTADT, N. Alcoholism and psychological differentiation: long-range effect of heavy drinking on field dependence. *Journal of Nervous and Mental Disease*, 1965, 140, 412-416.
- KARP, S., POSTER, D., & GOODMAN, A. Differentiation in alcoholic women. *Journal of Personality*, 1963, 31, 386.
- KARP, S., WITKIN, H., & GOODENOUGH, D. Alcoholism and psychological differentiation: Effect of alcohol on field dependence. *Journal of Abnormal Psychology*, 1965, 70, 262-265.
- MASLING, J., RABIE, L., & BLONDHEIM, S. H. Obesity, level of aspiration, and Rorschach and TAT measures of oral dependence. *Journal of Consulting Psychology*, 1967, 31, 233-239.
- MASLING, J., WEISS, L., & ROTHSCCHILD, B. Relationships of oral imagery to yielding behavior and birth order. *Journal of Consulting Psychology*, 1968, 32, 89-91.
- MCCORD, W., & MCCORD, J. *Origins of alcoholism*. Stanford: Stanford University Press, 1960.
- POLLNER, E. *Alcoholic personalities*. New York: Exposition Press, 1965.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
- WEINER, G. Neurotics', depressives' and alcoholics' oral Rorschach percepts. *Journal of Projective Techniques*, 1956, 20, 453-455.
- WITKIN, H., KARP, S., & GOODENOUGH, D. Dependency in alcoholics. *Quarterly Journal of Studies on Alcohol*, 1959, 20, 493-504.

(Received July 14, 1967)

⁴A friend of the second author, a physician with 20 yr. of experience in Israel, has reported that in all those years he has seen only one alcoholic—an American Marine.

HABITUATION OF THE ORIENTING RESPONSE AS A FUNCTION OF INDIVIDUAL DIFFERENCES IN ANXIETY AND AUTONOMIC LABILITY¹

EDWARD S. KATKIN AND ROBERT J. McCUBBIN

State University of New York at Buffalo

This study evaluated the role of individual differences in anxiety and autonomic lability on the habituation of GSR responses to auditory stimuli of moderate and very low intensity. Sixty *Ss* were used. They were divided into high- and low-anxious groups on the basis of Taylor *MA* scale scores and high and low autonomic lability groups on the basis of spontaneous GSR activity during a rest period. No relationship between Taylor score and GSR habituation was obtained for either stimulus intensity. Autonomically labile *Ss* showed no habituation to the moderate tone, but did habituate to the low intensity tone. Autonomically stable *Ss* habituated to both tones. It was concluded that individual differences in autonomic lability may predict whether an *S* will respond to moderate intensities with orienting or defensive reactions.

The purpose of this study was to evaluate the role of anxiety in the elicitation and habituation of orienting and defensive responses, two systems described by Sokolov (1960, 1963a, 1963b). The orienting response (OR), as Sokolov has conceptualized it, is a generalized response system possessing autonomic components, which is presumed to serve as a feedback mechanism facilitating peripheral receptor sensitivity. The defense reflex (DR), a second general response system, also possesses autonomic components, but it is distinguishable from the OR. According to Sokolov, ORs are elicited by low intensity "novel" stimuli, and DRs are elicited by high intensity, or noxious, varieties of stimulation. Furthermore, ORs habituate rapidly to repeated stimulation, whereas DRs do not and may even be intensified by repeated stimulation.

Although much research has been directed toward clarifying the effects of external stimulus properties on orienting behavior (Geer, 1966; Uno & Grings, 1965; Zimny & Schwabe, 1965), less attention has been given

to the effects of individual differences on the elicitation and habituation of ORs and DRs (Maltzman & Raskin, 1965). It is clear that an *E* can administer stimuli so intense that most *Ss* will judge them to be noxious; similarly, very weak stimuli can be administered so that few, if any, *Ss* will respond to them as if they were noxious. When moderately intense stimuli are presented, the influence of individual differences on response characteristics becomes crucial; some *Ss* are likely to respond as if the stimuli were quite intense, and other *Ss* are likely to respond as if they were not at all intense. Among the variety of individual difference variables of importance in contemporary personality research and theory, anxiety stands out as one which is related frequently to autonomic responsivity (see Martin, 1961). With respect to the issue at hand, it might be expected that anxious *Ss* would interpret moderately intense stimuli as noxious more frequently than nonanxious ones. Thus, anxious *Ss* would be expected to show autonomic responses characteristic of the DR to moderately intense stimuli, that is, responses which resist habituation; nonanxious *Ss* would be expected to show autonomic responses characteristic of the OR, that is, those which habituate rapidly. Therefore, the present experiment endeavored to investigate autonomic responses to moderately intense levels of auditory stimulation for *Ss* who differ in anxiety level.

¹The authors are indebted to S. Geisser for his aid in the statistical analyses. This research was supported in part by Grant No. MH-11989 from the National Institute of Mental Health, United States Public Health Service, and in part by a grant-in-aid from the Society of the Sigma Xi. Portions of this paper were presented by the second author at the meeting of the Eastern Psychological Association, Boston, April 1967.

As a comparison, stimuli of very low intensity were also administered, since it would be expected that such stimuli would elicit ORs from all Ss, irrespective of anxiety level.

Although anxiety is a construct which is presumed to have autonomic response components, much research reveals no relationship between clinically or psychometrically evaluated anxiety and autonomic response patterns (see Martin, 1961; Sarason, 1960). The clinical measures and psychometric scales used in most of this research are considered to be relatively valid indexes of anxiety, and the physiological measures with which they are compared are considered to reflect accurately autonomic nervous system (ANS) activity. Thus, the failure to obtain relationships between these indexes cannot be attributed easily to the invalidity of one or another of the measures employed. Rather, it may be that "anxiety" is not a unitary construct, but a complex one, composed at least of an autonomic component, reflected in ANS activity, and a "perceptual" or "cognitive" component, reflected in an individual's ability to recognize and report his ANS responses. Thus, one explanation for the noticeable lack of a relationship between self-report measures of anxiety and ANS activity may be simply that Ss who are aroused equally possess differential abilities to recognize and/or report the internal phenomena (Mandler, Mandler, & Uviller, 1958). The present experiment attempted to use both psychometric and physiological indexes as independent measures of anxiety.

Many psychometric instruments have earned their labels as "anxiety scales," but few physiological indexes have been so blessed. In the present investigation a measure of autonomic lability was used as an individual difference measure and employed as if it were an index of anxiety. The physiological measure employed, nonspecific (NS) electrodermal activity, is an index of spontaneous fluctuations in skin resistance which occur in the absence of external stimulation. NS activity has been shown to be an effective monotonic indicator of central nervous system activation (Greiner & Burch, 1955) and autonomic arousal (Silverman, Cohen, & Shmavonian, 1959), as well as a measure of

transitory response to stress (Katkin, 1965, 1966). Furthermore, Rappaport and Katkin (1967) have found that NS activity during moderately stressful situations is related to scores on the Taylor Manifest Anxiety (MA) scale (Taylor, 1953).

The paper-and-pencil measure employed in this study was the MA scale. This test is widely used and has been demonstrated to be related to clinical judgments of anxiety (Buss, 1955; Buss, Wiener, Durkee, & Baer, 1955; Eriksen & Davids, 1955; Hoyt & Magoon, 1954), as well as to other scalar measures of anxiety and general neuroticism (Bendig, 1957; Cattell & Scheier, 1958; Lauterbach, 1958). Furthermore, the MA scale has been described by its developers as an index of emotional responsivity, a hypothetical component of general drive (J. T. Spence & K. W. Spence, 1966; K. W. Spence, 1958). As such it has been compared directly with such concepts as "activation" and "arousal" (see Malmö, 1959).

The galvanic skin response (GSR) was chosen as the index of orienting and defensive behavior, and the following hypotheses were formulated: (a) Moderately intense stimuli will elicit larger GSRs than low intensity stimuli for all Ss, irrespective of anxiety level; (b) for moderately intense stimulation, nonanxious Ss will show rapid habituation of the GSR, typical of the OR, while anxious Ss will not show habituation, typical of the DR; and (c) for very low intensity stimuli, there will be no difference in rate of habituation of GSR for the two groups.

METHOD

Subjects

The MA scale and the MMPI L scale were administered to all male undergraduates in the introductory psychology course as part of a larger battery of personality tests. Subsequently, male Ss who scored below 7 on the L scale were recruited for the experiment. They were contacted by phone and asked to participate in an experiment concerning "psychophysiology." In this manner, 60 volunteers were recruited; they received one credit toward fulfilling a course requirement.

Apparatus

All Ss were run in a soundproof chamber equipped with a comfortable reclining lounge chair. A one-way

mirror behind the chair allowed *E* to monitor *Ss* visually from outside the chamber in order to observe possible artifact-producing movements. Skin resistance was detected by a pair of Beckman silver-silver chloride electrodes, employing Beckman NaCl electrode paste as a contact medium. The active electrode was placed on the left palm and the reference electrode on the left forearm. The electrodes were secured with stomasal collars and masking tape around the hand for additional support.

The amplification unit and polygraph were located outside the chamber. Skin resistance was amplified by a two-channel GSR amplifier, Model 201, manufactured by Bio-physical Research instruments. This amplifier, which with the 1 sq.cm. Beckman electrodes produced an effective current density of 20 microamp./sq.cm. for exosomatic recording, provided both high-gain and low-gain amplification of the skin-resistance signal. The high-gain amplification allowed for the recording of small amplitude NS activity, and the low-gain amplification allowed for the recording of larger amplitude electrodermal responses to the stimuli, and for the assessment of basal skin resistance levels. Both signals were recorded on a Grass polygraph, Model 5D.

Auditory stimulation was delivered to *Ss* through a headset by an Elco Audio Frequency Generator, Model 377. Stimulus duration and interstimulus intervals were controlled by a series of Hunter timers.

Procedure

Ss were initially assigned randomly to either a moderate intensity (M Int) or a low intensity (L Int) condition, but toward the end of the experiment randomness was abandoned in order to assure the assignment of 30 *Ss* to each group.

At the beginning of the experiment all *Ss* were told:

We are interested in how the autonomic nervous system reacts to stimulation. I would like you to sit here and relax while we measure your responses to some tones. You will hear them through these earphones, but first there will be a rest period during which you should try to relax. Be sure that you don't fall asleep, however.

The *E* then attached the electrodes to *Ss*, adjusted the headset for maximum comfort, and urged the *Ss* not to move their bodies, particularly their arms.

After a 10-min. adaptation period in which *Ss* acclimated to the situation, and resting levels of NS activity were recorded, stimulation began. All *Ss* received 15 presentations of a 1,000 cps sine-wave tone which lasted 2 sec. and was presented at variable intervals ranging from 30 to 90 sec., with a mean of 60 sec. For *Ss* in the M Int group, the volume level of the generator was set at 80; for *Ss* in the L Int group, the volume level was set at 20. These settings were chosen because previous experience suggested that they were easily distinguishable; the L Int tone was faint but audible, whereas the M Int tone was quite noticeable, but not noxious.

After 15 presentations of the stimuli, *Ss* were dismissed from the experiment. Data obtained from 1 *S* in the L Int condition were found to be unscorable, leaving 59 *Ss* in the final sample.

RESULTS

The GSR index of the OR was expressed as the change in low conductance, which was determined for each of the stimuli in the following manner: First, the prestimulus skin-resistance level was determined and designated R_1 ; then the skin-resistance level at the point of maximum deflection after stimulation was determined and designated R_2 . Response magnitude was then calculated by the formula $\log 1/R_2 - \log 1/R_1$. The 15 scores determined in this manner were then reduced to five trial blocks of three trials each.

The relationships among stimulus intensity, habituation across trial blocks, and scores on the *MA* scale will be discussed first. Then the relationships among stimulus intensity, habituation, and NS activity will be evaluated. Finally the relationship between *MA* scale scores and NS activity will be presented.

MA Scale Scores, Stimulus Intensity, and Habituation of GSR

Scores on the *MA* scale ranged from 4 to 35, with a median of 14. Those *Ss* who scored between 4 and 13 were designated low anxious (Lo A), and those who scored between 15 and 35 were designated high anxious (Hi A). Four *Ss* who scored at the median were eliminated. Thus, four experimental groups were constituted: M Int-Hi A ($N = 14$), M Int-Lo A ($N = 15$), L Int-Hi A ($N = 14$), and L Int-Lo A ($N = 12$). To provide equal cell frequencies for analysis of variance, 2 *Ss* were eliminated randomly from the M Int-Hi A group, 3 from the M Int-Lo A group, and 2 from the L Int-Hi A group. Thus the final data pool was based upon 48 *Ss*, divided equally among the four conditions.

The mean GSR amplitudes for *Ss* in the four experimental groups across five trial blocks are presented in Figure 1, in which it is evident that responses were greater in all trial blocks for *Ss* who received the moderate stimulation and that they diminished across trial blocks for all *Ss*, irrespective of

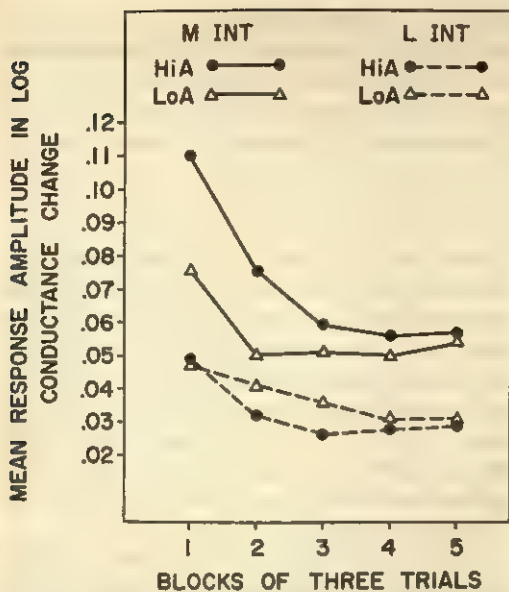


FIG. 1. Mean amplitude of response (change in log conductance) for Hi A and Lo A Ss in the M Int and L Int groups across five trial blocks.

anxiety level. These observations were confirmed by an analysis of variance (Type III; Lindquist, 1953) which yielded significant main effects for stimulus intensity ($F = 20.19$, $df = 1/44$, $p < .001$) and for trial blocks ($F = 18.00$, $df = 4/176$, $p < .001$). No main effect for MA scale scores and no interactions of MA scale scores with intensity or trial blocks were found.

NS Activity, Stimulus Intensity, and Habituation of the GSR

NS activity level was defined as the number of spontaneous fluctuations in skin resistance of at least 100 ohms magnitude occurring in the 2-min. period just prior to presentation of the first stimulus. This allowed Ss 8 min. in which to adjust to the experimental situation, thereby eliminating possible individual differences in NS activity which might have arisen from the newness of the situation. NS scores ranged from 0 to 34, with a median of 9. Twenty-five Ss whose NS responses ranged from 0 to 8 were designated the low NS (LNS) group, and 25 Ss whose responses ranged from 11 to 34 were designated the high NS (HNS) group. Five Ss whose NS levels were at the median and 4 Ss who had 10

responses were eliminated from the analysis. In this manner four groups of Ss were again constituted: M Int-HNS ($N = 15$), M Int-LNS ($N = 11$), L Int-HNS ($N = 10$), and L Int-LNS ($N = 14$). To provide equal cell frequencies for analysis of variance, 5 Ss were eliminated randomly from the M Int-HNS group, 1 from the M Int-LNS group, and 4 from the L Int-LNS group. Thus, the final analysis of these data was based upon a sample of 40 Ss, equally divided among the four experimental groups.

Mean change in log conductance for Ss in the four experimental groups across the five trial blocks is presented in Figure 2, which shows that the response amplitudes for Ss in the two M Int groups appear to be greater than comparable responses for Ss in the two L Int groups in all trial blocks. Figure 2 also indicates that the habituation rate for the M Int-LNS group appears steeper than that for the M Int-HNS group, and that the habituation rates for Ss in the two L Int groups are similar. Analysis of variance of these data (Type III) confirmed these observations, yielding a significant NS Activity \times Stimulus Intensity \times Trial Blocks interaction ($F = 3.80$, $df = 4/144$, $p < .01$), a significant Stimulus Intensity \times Trial Blocks interaction

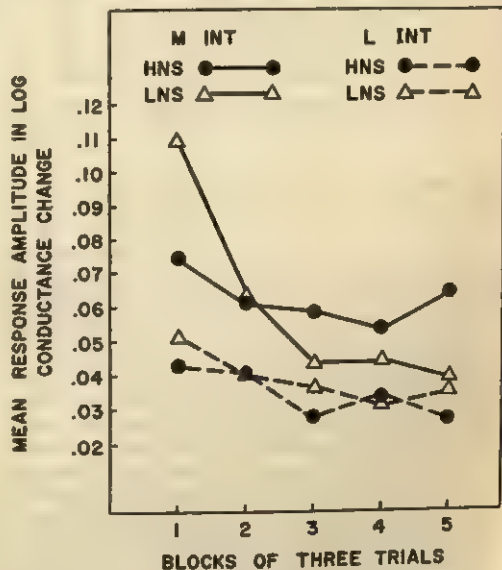


FIG. 2. Mean amplitude of response (change in log conductance) for HNS and LNS Ss in the M Int and L Int groups across five trial blocks.

($F = 3.45$, $df = 4/144$, $p < .025$), and a significant NS Activity \times Trial Blocks interaction ($F = 3.72$, $df = 4/144$, $p < .01$), as well as main effects for stimulus intensity ($F = 11.57$, $df = 1/36$, $p < .005$) and trial blocks ($F = 15.70$, $df = 4/144$, $p < .001$). It may also be noted that there were initial differences in response strength among the groups. Analysis of the first trial block indicated a significant difference between the two M Int groups and the two L Int groups ($F = 13.31$, $df = 1/36$, $p < .001$), but no differences between the HNS and LNS groups ($F = 3.31$, $df = 1/36$), and no interaction among the four groups ($F = 1.24$, $df = 1/36$). Thus, the obtained effects of NS activity were not attributable to initial differences in response magnitude for the HNS and LNS groups.

Subsequent analyses of the data confirmed that for Ss in the L Int conditions there was no differential effect of NS activity on habituation ($F = .58$, $df = 4/72$); GSRs habituated at the same rate irrespective of NS level ($F = 3.68$, $df = 4/72$, $p < .01$). However, for Ss in the M Int groups only, an analysis of variance confirmed the observation that habituation was steeper for the LNS group than for the HNS group ($F = 5.41$, $df = 4/72$, $p < .001$), and an additional analysis indicated that this interaction was determined by the fact that the M Int-HNS group showed no change across trial blocks ($F = .47$, $df = 4/72$), while the M Int-LNS group showed significant change across trial blocks ($F = 6.80$, $df = 4/72$, $p < .001$).²

Relationship between MA Scale Scores and NS Activity

To determine the relationship between the two measures of anxiety employed in this ex-

periment, a Pearson product-moment correlation was computed between NS activity levels and scores on the MA scale for all 59 Ss. The correlation obtained was low, negative, and not significantly different from a chance relationship ($r = -.17$).

DISCUSSION

In this experiment it was hypothesized that moderately intense stimuli would elicit larger autonomic responses than stimuli of very low intensity for all Ss regardless of anxiety level. It was also hypothesized that stimuli of very low intensity would elicit ORs for all Ss, but that moderately intense stimuli would elicit DRs for anxious Ss and ORs for nonanxious Ss. Anxiety was defined operationally by MA scale scores and NS activity.

The first hypothesis was confirmed, and the subsequent hypotheses were confirmed only for the autonomic measure of anxiety. Thus, when an index of autonomic lability was interpreted as a measure of anxiety, expectations about the relationship of anxiety to orienting behavior were sustained, but when a self-report index was used to define anxiety, predictions were not confirmed. Furthermore the self-report index and the measure of autonomic lability were not related.

The finding that scores on the MA scale were unrelated to magnitude and habituation of response, while disappointing, is consistent with previous findings which indicate that the MA scale and autonomic responses are independent (Katkin, 1965; Sarason, 1960). Thus, whereas the construct underlying MA scale scores has been related to clinical judgments of anxiety (Buss, 1955) and to conditionability (J. T. Spence & K. W. Spence, 1966), it has not been possible to demonstrate any consistent relationship between it and ANS measures.

The finding that autonomic responses typical of the DR were elicited from autonomically labile Ss and that responses characteristic of the OR were elicited from stable Ss confirmed the main hypotheses and enhanced the validity of NS activity level as a reliable index of individual differences in "anxiety." Support for the notion that anxiety level may contribute to the differential elicitation of ORs and DRs may be inferred also from Israel's

² In all of the reported analyses concerning main effects of trial blocks or any interactions of trial blocks with other variables there is a possibility inherent in repeated-measures designs of spuriously high F ratios resulting from inequalities and asymmetries of covariance matrices. In order to rule out this possibility, all degrees of freedom in tests involving trial blocks were adjusted according to the modification suggested by Greenhouse and Geisser (1959). Multiplying all df values by .57 (the epsilon value obtained from the covariance matrix) did not alter any of the conclusions derived from the analysis of variance, although in some cases alpha levels should be placed at the .03 level rather than at .01.

(1966) recent demonstration that schizophrenic patients show marked resistance to habituation of GSR, while normal controls show rapid habituation. Although Israel did not comment directly upon the anxiety levels of her Ss, it is commonly accepted that schizophrenics are more anxious than normals. Obviously, it would be of interest to investigate differences in levels of NS activity among psychiatric and nonpsychiatric patients. Further support for the present findings comes also from a recent report by Koepke and Pribram (1966), who found that in response to moderately intense tones GSRs of autonomically stabile Ss habituated and those of autonomically labile Ss did not. In their experiment, however, autonomic lability was measured by 1,000-ohm rather than 100-ohm deflections in skin resistance.³

Although the present findings suggest that NS activity may be a valid predictor of individual differences in autonomic responsivity, the generality of these results is limited. Clearly, the present contention that "anxiety" is being measured by this autonomic index requires further validation. To this end, future studies must be directed toward elucidating the relationships among NS activity and commonly accepted criteria for anxiety, such as clinical assessment, task performance, response to stress, and self-report (see Krause, 1961). Furthermore, although the present findings substantiate the hypotheses when this measure is conceptualized as an anxiety index, it is clear that the definition is problematic. It is obvious that it lacks certain traditional characteristics of anxiety. Most definitions of anxiety imply that an S's recognition of his unpleasant state is essential (Freud, 1936); yet Ss are notoriously unable to identify or describe their autonomic activity (Mandler et al., 1958; Rappaport & Katkin, 1967). It is possible, of course, that although Ss are unable to identify their NS activity they may be able to report reliably

upon aspects of themselves which are related to this activity. In that case, it would seem fruitful to attempt to delineate those characteristics of autonomically labile and stabile individuals which can be self-observed and self-reported.

To summarize, the present study suggested that individual differences in anxiety state will determine whether an S shows orienting or defensive response patterns to moderately intense stimuli. Finally, it was shown that an index of autonomic lability can be profitably used as a measure of individual differences in anxiety state and can predict Ss' characteristic mode of autonomic response, whereas a self-report index of anxiety cannot.

REFERENCES

- BENDIG, A. W. Extraversion, neuroticism, and manifest anxiety. *Journal of Consulting Psychology*, 1957, 21, 398.
- BUSS, A. H. A follow-up item analysis of the Taylor anxiety scale. *Journal of Clinical Psychology*, 1955, 11, 409-410.
- BUSS, A. H., WIENER, M., DURKEE, A., & BAER, M. The measurement of anxiety in clinical situations. *Journal of Consulting Psychology*, 1955, 19, 125-129.
- CATTELL, R. B., & SCHEIDER, I. H. The nature of anxiety: A review of thirteen multivariate analyses comprising 814 variables. *Psychological Reports*, 1958, 4, 351-388.
- ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the Hysteria-Psychasthenia scales from the MMPI. *Journal of Abnormal and Social Psychology*, 1955, 50, 135-137.
- FREUD, S. *The problem of anxiety*. New York: Norton, 1936.
- GEER, J. H. Effect of interstimulus intervals and rest-period length upon habituation of the orienting response. *Journal of Experimental Psychology*, 1966, 72, 617-619.
- GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, 1959, 24, 95-112.
- GREINER, T. H., & BURCH, N. R. Response of human GSR to drugs that influence the reticular formation of brain stem. *Federation Proceedings American Society of Experimental Biology*, 1955, 14, 346.
- HOYT, D. P., & MAGOON, T. M. A validation study of the Taylor Manifest Anxiety scale. *Journal of Clinical Psychology*, 1954, 10, 357-361.
- ISRAEL, N. R. Individual differences in GSR orienting response and cognitive control. *Journal of Experimental Research in Personality*, 1966, 1, 244-248.
- KATKIN, E. S. *The relationship between self-report and physiological indices of anxiety during differ-*

³There has been little research comparing the frequency of very small amplitude NS activity (100 ohms or less) with larger amplitude NS activity (500 ohms or more), but there is some evidence that the small amplitude responses are independent of basal conductance levels (Katkin, 1964), and therefore probably not affected as much by resting levels.

previous research in this area has indicated high reliability of judgments, but little validity. However, Kramer (1963) noted that:

The validity criteria, such as the Bernreuter Personality Inventory, which have been most frequently used in voice and personality studies, are often highly imperfect measures of those traits that they are used to validate [p. 248].

A recent study has supported the validity of the relationship between voice quality and at least one type of personality trait (Markel, Meisels, & Houck, 1964). In this study, rather than using a psychological test or inventory, speakers were actually either schizophrenic or nonschizophrenic. All speakers were hospital patients, and those classified as "schizophrenic" were patients on the psychiatric ward of the hospital and had been diagnosed schizophrenic by two psychiatrists. All speakers were recorded reading exactly the same paragraph, and the semantic differential was used to obtain ratings of the speakers' voices. College students using these ratings differentiated the voices of the schizophrenic and nonschizophrenic patients. The positive results of this study indicated that voice quality is, actually, a significant carrier of a "message" regarding personality and provided the impetus to develop a structural method to quantify voice qualities (Markel, 1965). The advantage of analyzing voice qualities by a structural method (e.g., pitch, loudness, tempo) rather than by a functional method (e.g., dominance, introversion, pleasant) is that a structural method can be applied meaningfully in any study in which voice quality is the behavioral variable. The present investigation employs this method to obtain voice-quality "profiles" and examines the personality traits of groups of speakers with different voice-quality profiles. This study is seen as a step in the direction of relating specific structural aspects of voice quality to specific "permanent" personality traits.

METHOD

Subjects

The Ss were 78 patients in the Veterans Administration Neuropsychiatric Hospital in Canadaiqua, New York. All were male, Caucasian, and spoke essentially the same dialect of English. Their age range was 20-58, with a mean age of 40.2.

The 78 Ss were selected from a total group of 90 patients who were admitted to the hospital from the time this study began. Twelve patients were not used as Ss in this study because of (a) their inability or refusal to provide an adequate speech sample, (b) a foreign accent, or (c) a dialect distinctly different from the other Ss. The 12 Ss were eliminated before voice qualities were analyzed and without any knowledge of the results of their psychological tests.

Personality Measure

The MMPI, administered to S as part of a battery of tests upon admission to the hospital, was given in booklet form. All 550 items were scored following the standard procedure described in the test manual. The Ss' T scores (*K* corrected) were obtained for the 10 clinical scales, and these T scores are the measures of personality structure used in this investigation.

Speech Samples

The speech samples were obtained as part of the admission procedure and at the same time the patients were being given other psychological tests. A microphone was put on lavalier fashion, and each patient was asked to read the sentence, "Many of the reports received last week showed a large increase in earnings," which was typed in capital letters. The same tape recorder (a Wollensak No.

TABLE 1
RELIABILITY OF RATING VOICE QUALITIES

Item	Pitch	Loudness	Tempo
Interrater reliability ^a			
Group 1			
Session 1	.79	.91	.91
Session 2 ^b	.82	.92	.90
Group 2			
Session 1	.72	.94	.88
Session 2 ^b	.72	.93	.87
Test-retest reliability ^c			
Group 1			
Sessions 1 and 2 ^d	.77*	.87*	.92*
Group 2			
Sessions 1 and 2 ^d	.79*	.95*	.92*
Group 1 and Group 2 ^e	.78*	.78*	.91*

Note.—*N* = 3 for all groups and sessions under "Interrater reliability."

^a Ebel (1951).

^b For Group 1, Rating Sessions 1 and 2 were separated by 4, 1, and 2 days for pitch, loudness, and tempo, respectively. For Group 2, Rating Session 2 followed immediately after Session 1.

^c Pearson *r*.

^d The means of the ratings for Ss in each session were the measures correlated.

^e The means of the ratings for Ss by each group in both sessions were the measures correlated.

* *p* < .001.

TABLE 2

MEAN T SCORES ON PITCH, LOUDNESS, AND TEMPO FOR THREE VOICE-QUALITY GROUPS

Group	N	Pitch	Loudness	Tempo
Peak-Pitch	29	56	43	46
Peak-Loudness	27	47	59	48
Peak-Tempo	22	45	47	58

1515) was used for each patient, the volume control was set at the same level for each patient, and during the period that these patients were being recorded the tape recorder was not moved or used for any other purpose. Following the collection of data, the speech samples of the 78 patients were dubbed in random order on a "test tape." The logic of this study requires the assumption that a single speech sample obtained from each patient was a typical one for that patient in a reading situation.

Rating Voice Qualities

In order to assure reliability, there were two groups of raters with three raters in each group. Group 1 consisted of three undergraduate students, and Group 2 consisted of three graduate students in the clinical psychology program at the University of Florida. None of the six raters had any previous training in rating voice qualities. They were trained to rate the qualities of pitch, loudness, and tempo following procedures which have been previously described (Markel, 1965). The rating scales for pitch, loudness, and tempo, respectively, were Low-High, Soft-Loud, and Slow-Fast. The scales were anchored by having a verbal label appear at each of the seven scale positions. For example, the scale for pitch was extremely Low, quite Low, slightly Low, average, slightly High, quite High, extremely High. The scale positions were also numbered from 1 to 7, and the raters used these numbers to indicate their ratings.

The speech samples on the training and practice tapes had been selected from previous studies so that no voice rated during the training procedures appeared on the test tapes of the 78 voices used in this study. None of the raters had any information as to the nature of the MMPI scores of Ss

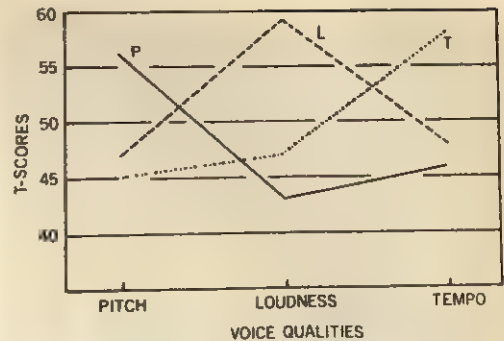


FIG. 1. Mean *T* scores on pitch, loudness, and tempo for the three voice-quality groups: P = Peak-Pitch; L = Peak-Loudness; T = Peak-Tempo.

whose voices they were rating. Both groups rated each voice quality twice. For Group 1, Rating Sessions 1 and 2 were separated by 4, 1, and 2 days for pitch, loudness, and tempo, respectively. For Group 2, Rating Session 2 followed immediately after Rating Session 1 for each voice quality. Interrater reliability for each group at each rating session was estimated by the method developed by Ebel (1951) and specifically followed the procedures described by Guilford (1954). These reliabilities are listed in Table 1. Test-retest reliability was determined by the correlation between the mean ratings of the 78 Ss. For the between-rating session correlation for each group of raters, the scores for each *S* were means of the three ratings by that group in each rating session. For the between-groups correlation the scores of each *S* were the means of six ratings by the three judges in a group in the two rating sessions. These correlations are also listed in Table 1.

Voice-Quality Profiles

The mean of the 12 ratings given a particular *S* on a particular voice quality was taken as that *S*'s score on that voice quality. In order to compare the scores for pitch, loudness, and tempo, the scores for the 78 Ss on each quality were transformed to *T* scores ($\bar{X} = 50$, $SD = 10$). *T* scores on the three parameters were then compared, and each *S* was assigned to one of three groups according to which

TABLE 3

MEAN T SCORES ON MMPI SCALES FOR THE THREE VOICE-QUALITY GROUPS

Group	N	Validity scale			Clinical scale										\bar{X}
		L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	
Peak-Pitch	29	50	64	51	65	79	65	69	59	62	72	71	60	61	66
Peak-Loudness	27	51	63	53	72	74	72	72	60	65	72	73	59	55	67
Peak-Tempo	22	51	57	53	56	68	63	70	60	61	62	62	61	51	61
\bar{X}					65	74	67	71	60	62	69	69	60	56	

one of the three voice qualities had attained the highest T score. The groups were labeled Peak-Pitch, Peak-Loudness, and Peak-Tempo. The mean voice-quality T scores and the N in each of the three profile groups are listed in Table 2 and plotted in Figure 1. One-way analysis of variance was used to evaluate the significance of the difference between the groups for each voice quality, and for each of the three voice qualities the difference between groups was significant beyond the .01 level, $F(2/75) = 10.62, 37.89, 14.76$, for pitch, loudness, and tempo, respectively. The Newman-Keuls method (Winer, 1962, p. 309) was used to examine the significance of the difference between the means of the three groups for each quality. The results indicate that for each voice quality the group whose peak was on that quality was significantly different from the other two groups, but the other two groups did not differ significantly from one another on that voice quality.

The mean ages for the three groups were 38, 43, and 40 for Peak-Pitch, Peak-Loudness, and Peak-Tempo, respectively. Analysis of variance indicated the differences in age among the groups were non-significant.

RESULTS¹

The significance of the difference between the 10 MMPI clinical scales for the three voice-quality profile groups was evaluated by a 3×10 analysis of variance for repeated measures (Winer, 1962, pp. 374-377). The mean T scores for the three voice-quality groups for each of the 10 clinical scales are listed in Table 3 and plotted in Figure 2.

¹The author is grateful to Jennie L. Grossman, associate in research at the University of Florida Computing Center, for her assistance in analyzing the data.

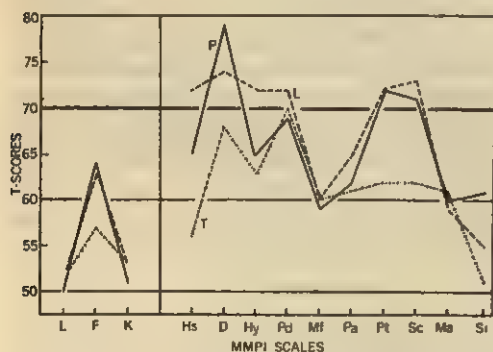


FIG. 2. Mean T scores on the MMPI scales for the three voice-quality groups. (Note: Analysis of the difference between the groups is based only on the 10 clinical scales.)

TABLE 4
ANALYSIS OF VARIANCE SUMMARY

Source of Variation	SS	df	MS	F
Between Ss		77		
Voice-quality groups (A)	4,798	2	2,399	2.29
Subjects (W)	78,652	75	1,049	
Within Ss		702		
MMPI scales (B)	23,119	9	2,569	21.35*
A \times B	5,351	18	297	2.47*
B \times W	81,212	675	120	

* $p < .01$.

(However, note that Table 3 and Figure 2 include the group means on the validity scales L , F , and K , but these scales were not included in the analysis of variance.) The results of the analysis of variance were as follows: Factor A (voice-quality groups), ns ; Factor B (MMPI scales), $F = 21.35$, $df = 9/675$, $p < .01$; $A \times B$, $F = 2.47$, $df = 18/675$, $p < .01$. The analysis of variance

TABLE 5
ANALYSIS OF VARIANCE SUMMARIES

Scale	SS	df	MS	F
<i>Hs</i>				
Between Ss	2,840	2	1,420	6.44**
Within Ss	16,542	75	221	
<i>D</i>				
Between Ss	1,399	2	699	1.98
Within Ss	26,263	75	350	
<i>Hy</i>				
Between Ss	1,124	2	562	3.85*
Within Ss	10,956	75	146	
<i>Pd</i>				
Between Ss	139	2	69	0.45
Within Ss	11,667	75	156	
<i>Ma</i>				
Between Ss	19	2	9	0.07
Within Ss	9,590	75	128	
<i>Pa</i>				
Between Ss	230	2	115	0.67
Within Ss	12,816	75	171	
<i>Pt</i>				
Between Ss	1,474	2	737	2.55
Within Ss	21,692	75	289	
<i>Sc</i>				
Between Ss	1,559	2	779	2.29
Within Ss	25,510	75	340	
<i>Si</i>				
Between Ss	51	2	26	0.14
Within Ss	13,936	75	186	
<i>Si</i>				
Between Ss	1,313	2	657	4.52*
Within Ss	10,892	75	145	

* $p < .05$.

** $p < .01$.

terms are summarized in Table 4. Inspection of Figure 2 indicates how the different "shapes" of the MMPI profiles for the three voice-quality groups produced the significant interaction effect.

Simple analysis of variance was employed to examine the difference between the groups on each of the 10 clinical scales. The analysis of variance terms are summarized in Table 5. This table indicates significant differences between the groups on three scales: *Hs* ($p < .01$), *Hy* ($p < .05$), and *Si* ($p < .05$).

DISCUSSION

The most important finding of this study is that the mean MMPI profiles for the three voice-quality groups are significantly different. Following the coding rules developed by Marks and Seeman (1963), a "best fit" of code type for the mean profiles of the Peak-Pitch and Peak-Loudness groups is 2-8 and 3-2-1, respectively. The Marks and Seeman actuarial information for code types 2-8 and 3-2-1 indicates a possible direction of the relationship personality traits and voice-quality profiles. In actuarial terms, Marks and Seeman (1963) stated an individual with a 2-8 profile is likely to be diagnosed as psychotic (p. 138), have a poor prognosis (p. 142), and be described as, "Keeps people at a distance; avoids close interpersonal relationships" (p. 138). On the other hand, an individual with a 3-2-1 profile is likely to be diagnosed as psychoneurotic (p. 152), have a good prognosis (p. 156), and be described as, "exhibits depression (manifest sad mood)" (p. 152). The profile of the Peak-Tempo group, with only *Pd* elevated, is unclassifiable as to code type.

Another indication of the relationship between personality traits and voice-quality profiles is seen in the significant statistical differences between the voice-quality profile groups of Scales *Hs*, *Hy*, and *Si*. Marks and Seeman (1963, pp. 307-310) provided a chart of actuarial statements regarding various levels of elevation on single scales.² The

following actuarial statements are taken from this chart and are based on the mean *T* scores for the voice-quality profile groups on Scales *Hs*, *Hy*, and *Si*.

Peak-Pitch: (*Hs*) Slightly more than average number of physical complaints. Some concern about bodily functions and physical health. (*Hy*) Normal level of hysterical symptoms. (*Si*) Probably reserved in unfamiliar social situations.

Peak-Loudness: (*Hs*) Considerable number of physical complaints. Prominent concern with bodily functions. (*Hy*) Probably somewhat immature, egocentric, suggestible, and demanding. (*Si*) Has capacity to maintain adequate social relationships.

Peak-Tempo: (*Hs*) Number of physical symptoms and concern with bodily functions fairly typical for clinic patients. (*Hy*) Normal level of hysterical symptoms. (*Si*) Has capacity to maintain adequate social relationships.

Since the actuarial statements regarding both profile differences and individual scale differences are based on mean scale scores, these statements must be taken as "first approximations" of the specific personality traits related to specific voice-quality profiles. However, the statistical results clearly support the hypothesis that the three voice-quality profile groups represent three different personality types, and these results are a strong indication that the voice-quality profile is a significant behavioral variable.

REFERENCES

- ABRAHAM, K. Psycho-analytical studies on character-formation. In *Selected papers of Karl Abraham*. New York: Basic Books, 1953.
- EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407-424.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- KRAMER, E. Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin*, 1963, 60, 408-420.
- MAHL, G. F., & SCHULZE, G. Psychological research in the extralinguistic area. In T. A. Sebeok, A. S. Hayes, & M. C. Bateson (Eds.), *Approaches to semiotics*. The Hague: Mouton, 1964.
- MARKEL, N. N. The reliability of coding para-language: Pitch, loudness, and tempo. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 306-308.

² Marks and Seeman (1963, p. 307) attributed this chart to Wendell M. Swenson. No specific statement is provided in this chart for the normal level of elevation on Scale *Hy*.

- MARKEL, N. N., MEISELS M., & HOUCK, J. E. Judging personality from voice quality. *Journal of Abnormal and Social Psychology*, 1964, 69, 458-463.
- MARKS, P. A., & SEEMAN, W. *Actuarial description of abnormal personality*. Baltimore, Md.: Williams & Wilkins, 1963.
- OSGOOD, C. E., & SEBEOK, T. A. Psycholinguistics, a survey of theory and research problems. *Journal of Abnormal and Social Psychology*, 1954, 49 (4, Pt. 2).
- REICH, W. On character analysis. In R. Fliess (Ed.), *The psychoanalytic reader*. Vol. 1. New York: International Universities Press, 1948.
- SAPIR, E. Speech as a personality trait. *American Journal of Sociology*, 1927, 32, 892-905.
- SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 17, 1967)

A REEXAMINATION OF THE RELATIONSHIP BETWEEN BODY BOUNDARY AND SITE OF DISEASE¹

MICHAEL HIRT

AND

RICHARD KURTZ²

Kent State University

University of Cincinnati College of Medicine

The relationship between body-boundary perception and body site of disease was examined. In Experiment I, a body-interior group (10 asthmatics plus 10 cardiacs) was compared with a body-exterior group (10 arthritics plus 10 dermatologic patients). In Experiment II Ss were more carefully screened in terms of their medical histories. Measures of psychopathology from the Holtzman Inkblot Test and the MMPI were used as covariants in both experiments. The results failed to replicate previously reported empirical findings and challenged the construct validity of the Fisher and Cleveland model.

Numerous studies (for a recent summary, see Fisher, 1963), based upon the model originally developed by Fisher and Cleveland (1958), have attempted to demonstrate a relationship between the degree of definiteness ascribed to one's body boundary and diseases with various psychosomatic symptoms. This assumed relationship is based upon the assumption that body image is influential in determining the body site at which symptoms occur whenever stress causes psychosomatic illness. The additional assumption is made that psychosomatic illness represents an extreme manifestation of underlying patterns of physiological reactivity. The Fisher and Cleveland model is based upon the hypothesis that individuals with definite body-image boundaries are characterized by relatively low reactivity in the body interior (heart, lungs).

Studies which have examined this relationship have inferred body-image boundaries from responses to Rorschach inkblots. Those Rorschach responses which were protective, covering, or containing (e.g., man in armor, turtle with a shell, mummy wrapped up) were scored as Barrier responses. The greater the Barrier score, the more definite was the inferred body-boundary perception. Relatively high Barrier scores were reported more frequently among individuals with body-exterior reactions (e.g., neurodermatitis or functional paralysis of musculature).

Penetration Rorschach responses are those in which there is a violation, disruption, or passage through a boundary region (e.g., crushed bug, bullet entering flesh, worn-out flower, vagina). The greater the Penetration score, the more indefinite the body-boundary perception was inferred to be. A relatively larger number of Penetration responses were reported to occur in individuals with body-interior reactions (e.g., stomach ulcers, bronchial asthma, cardiac disease). Because of the following limitations, repeated in practically all the work of these authors (Mednick, 1959), it is extremely difficult to assess the adequacy of this body-image theory:

1. Failure to control for variables which are important covariants of Penetration and Barrier scores. Barrier scores, for instance, appear clearly related to $F+\%$ and probably to other Rorschach indicators of "ego strength" or degree of psychopathology.

2. The arbitrary development of a scoring system for the Rorschach without adequate supporting data relevant to its reliability or validity.

3. The consistent comparison of only *one* body-interior disease group (e.g., stomach ulcers) against *one* body-exterior disease group (e.g., rheumatoid arthritis). If this construct is valid, then differences in body-boundary perception should manifest themselves across multiple groups of body-interior and body-exterior diseases; differences between two specific groups might be due to other factors (e.g., differences in psychopathology, chronicity of the disease, or degree of disability).

¹ A modified version of this paper was presented at the meeting of the American Psychological Association, Washington, D. C., September 1967.

² Now at Washington University.

The purpose of this study is to consider these difficulties and to provide a more rigorous test of the hypothesis that individuals with body-exterior reactions perceive their body boundary as more definite than those individuals with body-interior reactions.

METHOD

All Ss were tested with the Holtzman Inkblot Test (Holtzman, Thorpe, Swartz, & Herron, 1961) and the MMPI. While providing psychometric improvements over the Rorschach procedure, the Holtzman inkblots, at least with respect to Penetration and Barrier scores, are almost identical, qualitatively, to their Rorschach counterparts.

The Holtzman inkblots were administered and scored by an examiner who was unaware of the purpose of the study. Protocols were scored for Penetration, Barrier, and Integration. This last variable is the single most sensitive Holtzman index of psychopathology and was examined as one of the covariants in the analysis of the data.

The study was divided into two parts. In Experiment I, Ss were selected exclusively on the basis of their current hospital diagnosis. In the literature reported there is usually insufficient information provided on S selection. It is assumed here that previous studies assigned Ss to the body-interior and body-exterior groups on some basis similar to current admitting diagnosis. The body-interior group consisted of two subgroups: 10 Ss with cardiac pathology and 10 Ss with bronchial asthma. The body-exterior group also consisted of two subgroups: 10 dermatologic patients and 10 Ss with rheumatoid arthritis.

The data analysis consisted of examining the combined body-interior groups (asthmatics, cardiacs) and combined body-exterior groups (dermatologic patients, arthritics) for differences in Penetration and Barrier responses. Additional analysis consisted of comparing Ss for Penetration and Barrier responses while examining the role of psychopathology as a covariant. In addition to the Holtzman variable of Integration, a psychopathology index from the MMPI was used. On the basis of Goldberg's (1965) recent evaluation, the simple linear composite of scales $L + Pa + Sc - Hy - Pt$ provides the single most valid MMPI index of psychoticism versus neuroticism. This index appears to be a continuum, ranging from "normal" to neurotic to psychotic. It seemed desirable to examine an index of psychopathology which did not overlap with the Holtzman variables.

Experiment II was designed to determine whether the previously reported relationships between body boundary and site of body reaction would reappear if S selection was more stringent. Accordingly, the charts of numerous hospitalized patients were examined by a physician experienced in both psychiatry and internal medicine. Three groups of eight patients each were thus selected. One group was a "pure" body-interior disease group in which all of

the present and past recorded medical problems of each S were consistent with a body-interior site. The second group was a "pure" body-exterior one. The third group of patients was mixed, composed of patients whose medical histories indicated that they had in the past reacted, or were currently reacting, with both body-interior and body-exterior sites. This assignment of patients to various diagnostic groups was also done independently and without knowledge of Ss' Holtzman protocols.

In Experiment II the data analysis was essentially the same as that for Experiment I. An additional refinement was introduced through the use of two groups which were "pure" and one "mixed" with respect to body site of disease. It was thus assumed that differences in body-boundary perception would form a continuum which could be inferred from three rather than two points.

RESULTS

The mean scores on Penetration, Barrier, Integration, and MMPI index of psychopathology for each of the four groups of Ss are presented in Table 1. It can be seen from this table that Ss with a body-interior site of disease differ from Ss with a body-exterior site of disease on the Penetration variable and on the MMPI index of psychopathology. The direction of the relationship between site of disease and Penetration responses was inconsistent with Fisher and Cleveland's (1958) theoretical model, however.

Specifically, comparison of the groups with body-interior diseases (asthma, cardiac) with the two groups with body-exterior diseases (dermatology, arthritis) yielded chance differences for both Penetration and Barrier responses. Use of Integration as a covariant yielded results which approached statistical significance, but in a direction opposite to

TABLE 1
MEAN SCORES ON PENETRATION, BARRIER, INTEGRATION, AND MMPI INDEX OF PSYCHOPATHOLOGY: EXPERIMENT I

Variable	Subjects					
	Body interior			Body exterior		
	Heart ^a	Asthma ^a	Total	Arthritis ^a	Dermatitis ^a	Total
Penetration	3.3	3.2	3.2	2.1	8.7	5.4
Barrier	5.0	4.5	4.8	3.0	7.2	5.1
Integration	12.4	4.3	8.4	8.1	11.0	9.6
MMPI index	36.9	36.3	36.6	46.8	48.3	47.5

^a N = 10.

TABLE 2

ANALYSIS OF VARIANCE AND COVARIANCE FOR PENETRATION, INTEGRATION, AND MMPI INDEX:
EXPERIMENT I

Source	SS	df	MS	F
Total	826.78	39		
Error	780.55	38	20.54	
Treatments	46.22	1	46.22	2.25
Adjusted				
Integration				
Total	788.22	38		
Error	750.50	37		
Treatments	38.02	1	38.02	1.87
Adjusted				
MMPI index				
Total	805.37	38		
Error	774.88	37	20.94	
Treatments	30.49	1	30.49	1.46

that predicted by the theory. That is, the group with the "interior" body site of disease gave fewer Penetration and more Barrier responses. Use of the MMPI index of psychopathology as the covariant also failed to differentiate the groups. These analyses are summarized in Tables 2 and 3.

In Experiment II the same data analysis was carried out with three groups of very carefully selected Ss. Two of these groups were very homogeneous with respect to having either an interior or exterior body site of disease; the third group included Ss with both types of diseases. These analyses are summarized in Tables 4 and 5. It can be seen from these tables that even these carefully

TABLE 3

ANALYSIS OF VARIANCE AND COVARIANCE FOR BARRIER, INTEGRATION, AND MMPI INDEX:
EXPERIMENT I

Source	SS	df	MS	F
Total	584.78	39		
Error	583.55	38	15.36	
Treatments	1.22	1	1.22	
Adjusted				
Integration				
Total	465.64	38		
Error	465.63	37	12.58	
Treatments	.005	1	.005	
Adjusted				
MMPI index				
Total	549.87	38		
Error	538.24	37	14.55	
Treatments	11.63	1	11.63	1.25

TABLE 4

ANALYSIS OF VARIANCE AND COVARIANCE FOR PENETRATION, INTEGRATION, AND MMPI INDEX:
EXPERIMENT II

Source	SS	df	MS	F
Total	263.33	23		
Error	217.25	21	10.34	
Treatments	46.08	2	23.04	2.23
Adjusted				
Integration				
Total	248.76	22		
Error	171.41	20	8.57	
Treatments	77.35	2	38.67	4.51*
Adjusted				
MMPI index				
Total	674.94	22		
Error	655.81	20	32.79	
Treatments	19.13	2	9.57	

* $p < .05$.

selected Ss did not differ significantly in the frequency of their Barrier and Penetration responses. Penetration-response differences approached statistical significance, but in the direction opposite to that predicted by the theory. When Integration scores were used as a covariant, these differences were statistically significant ($p < .05$).

In order to better understand the role of psychopathology as a covariant of Penetration and Barrier responses, the product-moment correlations between these two Holtzman inkblot variables and the MMPI index of psychopathology were computed. For the group with the interior body site of disease,

TABLE 5

ANALYSIS OF VARIANCE AND COVARIANCE FOR BARRIER, INTEGRATION, AND MMPI INDEX:
EXPERIMENT II

Source	SS	df	MS	F
Total	386.50	23		
Error	385.50	21	18.36	
Treatments	1.00	2	.50	
Adjusted				
Integration				
Total	305.42	22		
Error	287.43	20	14.37	
Treatments	17.99	2	9.00	
Adjusted				
MMPI index				
Total	390.70	22		
Error	385.70	20	19.29	
Treatments	5.00	2	2.50	

Penetration correlated $-.626$ with the MMPI index and Barrier correlated $-.451$. Both of these correlations are statistically significant and of "respectable" magnitude. The correlations between the MMPI index and Penetration and Barrier scores for the group with the exterior body site of disease were $.315$ and $-.159$, respectively. Neither is statistically significant. These findings suggest that the relationships between Barrier and Penetration responses and the index of psychopathology used are of a chance nature among Ss with an exterior body site of disease, though they are significant for patients with interior-disease body sites. Experiment II yielded similar findings; the correlations between the MMPI index of psychopathology and Barrier and Penetration responses were $-.475$ and $-.639$, respectively, for the group with the interior-disease body site, and $-.475$ and $.162$ for the group with the exterior body site of disease.

The inverse relationship between the MMPI index of psychopathology and Penetration responses is surprising. The MMPI index achieved greatest accuracy in classifying patients when scores of 49 or higher were classified as psychotic and scores below 40 were classified as neurotic. As Penetration responses appear qualitatively pathological, a positive relationship between these two variables would have appeared more understandable.

DISCUSSION

This study has examined the theoretical construct postulated by Fisher and Cleveland (1958) which relates perception of body boundary with site of disease. The empirical data previously reported have provided only tangential support for such a link. Aside from methodological limitations, results have been reported only from the comparison of two groups at one time. A more critical test of the construct validity of the psychosomatic relationship postulated would necessitate a comparison between several groups with body-interior disease sites and several groups with body-exterior disease sites. In the present study, where four such groups were compared, there were no differences between the groups in body-boundary perception.

Considering alternate explanations for the

positive results previously reported, two possibilities suggest themselves; one is *S* selection, and the second is uncontrolled sources of variance (i.e., other than body-boundary perception).

As to *S* selection, the present authors were impressed by the difficulty in finding Ss whose medical histories clearly reflected that their body disease sites had always been internal or external. Many Ss hospitalized with a diagnosis of a disease having a given body site were also found to have histories of diseases with the opposite body site. After examining many such records, *S* samples with diseases whose body sites were "pure," in the sense that they were all either external or internal, were found. The body-boundary perception of these groups either was not significantly different or differed in a direction opposite to that predicted by the theory.

Such a finding suggests the need to examine sources of variance other than body-boundary perception. In this study, one such source, psychopathology, was examined. When an index of psychopathology from the Holtzman inkblots was used as the covariant, the groups did show significant differences in body-boundary perception, but in a direction opposite to that postulated by Fisher and Cleveland. The present study thus raises the possibility that the differences previously reported between single groups could have been due to differences in psychopathology, rather than to body-boundary perception.

Neither the construct validity nor the predictive validity of the Fisher-Cleveland model appears to be confirmed.

REFERENCES

- FISHER, S. A further appraisal of the body boundary concept. *Journal of Consulting Psychology*, 1963, 27, 62-74.
- FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton, N. J.: Van Nostrand, 1958.
- GOLDBERG, L. R. Diagnosticians versus diagnostic signs: The diagnosis of psychosis versus neurosis from the MMPI. *Psychological Monographs*, 1965, 79(9, Whole No. 602).
- HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality*. Austin: University of Texas Press, 1961.
- MEDNICK, S. A. The body's barriers go Rorschach. *Contemporary Psychology*, 1959, 4, 276-277.

(Received July 21, 1967)

AN INVESTIGATION OF METHODS OF PRESENTING INCOMPLETE SENTENCE STIMULI¹

FRANK A. WOOD²

College of Education, Ohio State University

One hundred and forty-four coeds completed two halves of a modified Rotter Incomplete Sentences Blank (ISB) with the Marlowe-Crowne Social Desirability (*SD*) scale interpolated. Modified ISB halves were constructed using personal pronouns (P), proper names (I), or neither (S). The instructions systematically varied between conventional and verbal speed test type. No significant differences were found using ISB difference scores or *SD* scale raw scores. Using ISB raw scores, differences appeared due to forms but not order or instructions. Form means were significantly different in ascending order S, P, I. These findings were cross-validated and interpreted as showing greater projection with disguised formats but not instructions. Intercorrelations were discussed, including negative correlations of the *SD* scale with certain ISB forms.

The complex interaction of stimulus material and its presentation with the motivations of the *S* constitute an inadequately understood phenomenon at best. The three variables—stimulus, personality, and environment—are usually found represented in practice as the test, *S*, and the instructions to *S*, respectively.

In the monograph by Getzels and Walsh (1958) on the PDPQ technique, or the method of Paired Direct and Projective Questionnaires, students were presented first with a sentence-completion test under the guise of a verbal speed test. The sentence stems had been altered with boys' and girls' proper names replacing personal pronouns ("I sometimes think my father is. . . ." became "Ruth sometimes thought her father was. . . ."), and the order of sentences was changed. Two weeks later the students completed the unaltered sentence-completion test according to the usual test instructions. Responses on both forms were scored reliably as positive, neutral, or negative. The verbal speed test form was found to elicit a significantly higher score (more negatives) consistently across grade

levels. It was felt that this indicated less guarding and more projection with the impersonal form than with the personal form of the PDPQ.

This PDPQ study left many questions unanswered. How much of the greater projection of the verbal speed test form is due to the press of speed and how much to the impersonal nature of the sentence stems? How robust is the effect? Would reversing the order clue the students so that they would guard on the impersonal form as well? And what would happen in the case of *nonpersonal* stems (such as, "Sometimes. . . ." or "The only trouble. . . .")? Little research is available which has attempted to answer these questions directly (cf. Cromwell & Lundy, 1954; Ellis, 1947; Meltzoff, 1951; Sacks, 1949; Stein, 1947).

Another consideration basic to the current research was the sensitivity of the particular type of sentence-completion test developed by Getzels and Walsh (1958). It was felt that sentence stems which were not as specific (not as long) offered greater projective possibilities (Rotter & Rafferty, 1950). A scoring system more rigorous than the positive-neutral-negative trichotomy of Getzels and Walsh also was suggested as desirable. Further, it was speculated that empirical scoring criteria would be more sensitive than criteria derived rationally. Thus, it was decided that the Rotter Incomplete Sentences Blank, College Form (ISB) would be used as the basic instrument in this research. The ISB was convenient in

¹ This paper is based upon the author's dissertation, submitted in partial fulfillment of the requirements for the PhD degree at Purdue University. The author wishes to thank James D. Linden, committee chairman, for his assistance. Special thanks is due Rhoda Wood, the author's wife, for her invaluable assistance in scoring the tests and computing results.

² Now at St. Lawrence University.

that, with virtually no change of sentence stems, it easily was divided into two 20-sentence halves—one-half with only personal pronoun (P) stems and one-half with only standard, nonpersonal (S) stems. A third, impersonal (I) form was constructed by replacing the personal pronouns in the P sentence stems with random proper female names.

A third consideration basic to this research was that Getzels and Walsh also discriminated between a personal hypothesis (set) and the reaction finally expressed by the individual. They believed that the PDPQ technique provided a measure of this distinction. The Direct Questionnaire, being in self-report form, elicited the censored, expressed reaction, while the Projective Questionnaire, "containing exactly the same items phrased in the third person," elicited the projectively revealed, personal hypothesis. It was reasonable to conjecture that any such distinction between the personal hypothesis and the expressed reaction might have evolved through the social motives of the individual, that is, by social guarding or the tendency of the individual to respond in a socially desirable manner. Accordingly, it was felt appropriate to obtain for each *S* in this current study a measure of this social desirability tendency. The Marlowe-Crowne Social Desirability (*SD*) scale was chosen as the instrument to ascertain how much, relatively, each *S* was motivated to respond in a socially desirable manner.

The experimental design, as established, included as independent variables the form of the stimuli (P, S, or I), the mode of stimuli presentation (timed, "verbal speed test" instructions, T; or untimed, "personality test" instructions, U), and the orders of the presentation of the form and modes to *Ss*. Accordingly, the following three hypotheses were tested, and the significant findings were cross-validated:

1. There will be no significant difference in scores due to the different forms of stimuli presented.
2. There will be no significant difference in scores due to the different modes of presenting the stimuli.
3. There will be no significant difference in

scores due to the order of presenting the forms and/or modes.

METHOD

The three forms (P, S, and I) were each mimeographed with two separate sets of instructions: T and U. Thus, there were six possible combinations of different tests and instructions: PT, PU, ST, SU, IT, and IU.

The *Ss* used in the first phase of the experiment were 144 female undergraduate students at Purdue University. The *Ss* used in the second phase were 32 female undergraduate volunteers at Ohio State University. Demographic information descriptive of each *S* group corresponded closely.

Each *S* in the first phase of the experiment completed one pair of sentence blanks from among the 36 pairs of sentence blanks possible. For practical purposes only one set of instructions (T or U) at a given time could be administered to a given group of *Ss*. The instructions could differ between the first and the second blank administered to each group, however. The form of the test (P, S, or I) was randomly assigned to *S* within each administration of each blank. Between the administration of the first and the second blank, each *S* completed the *SD* scale. (Neither the instructions nor the form of the *SD* scale ever varied.)

Each *S* in the second phase of the experiment completed all three sentence blanks, given in the order I, S, P. The blanks were administered to as many as 10 *Ss* at any one time under U instructions.

RESULTS AND DISCUSSION

The reliability coefficients obtained were Form P, .84; Form S, .74; Form I, .73. All three null hypotheses were accepted when either ISB difference-score data or Marlowe-Crowne raw-score data were used (see Table 1). However, using the ISB raw-score data, the first null hypothesis, that there would be no difference in score due to the form of the test, was rejected. The other null hypothesis concerning the possible effects of instructions

TABLE 1
F VALUES FOR ANALYSES OF VARIANCE
USING DIFFERENT INPUT DATA

Source	Analysis I	Analysis II	Analysis III
Form (F)	2.41	0.09	58.90*
Order (O)	0.00	0.00	0.12
Instructions (I)	0.60	0.09	2.00
F × O	2.68	0.46	2.00
F × I	0.05	0.80	0.71
O × I	0.58	3.23	0.61
F × O × I	0.20	1.16	0.75

* $p \leq .05$.

or order of presentation was accepted. (The level of significance held critical for all tests made for all data was defined as .05.)

The ISB thus was altered in certain ways without effecting a significant difference in the scores obtained. Likewise, the *SD* scale showed itself to be unaffected by the form, instructions, or order of presentation of the ISB. However, differences in the form of the ISB did produce a significant difference in ISB raw scores. It can be surmised that the differences in test content, as opposed to the order of presentation or the instructions, were largely responsible for the score differences observed previously with the PDPQ. The necessity noted by Getzels and Walsh (1958) of administering the impersonal (proper name) form first so as not to give *Ss* a clue does not appear to be a real consideration here.

In view of the significant difference obtained due to form, a Newman-Keuls test for ordered mean differences (Winer, 1962) was performed to ascertain what specific form-score differences were obtained on the ISB. It was found that each form was significantly different from every other form in the ascending order *S*, *P*, *I* (see Table 2). Form *I* (which was constructed as a proper-name variant of Form *P*) tended to elicit responses that were scored higher, that is, more maladjusted, according to the Rotter criteria. Form *S*, which contained the totally nonpersonal Rotter items, produced the *least* maladjusted responses.

This difference due to the form of the test was further investigated in the second phase of the experiment, where every *S* took all three forms. A *t* test revealed that the means

TABLE 2
A COMPARISON OF FORM MEAN SCORES FOR INITIAL AND CROSS-VALIDATION SAMPLES

Form	Initial sample		Cross-validation samples	
	<i>N</i>	<i>M</i>	<i>N</i>	<i>M</i>
<i>S</i>	96	57.4	32	57.4
<i>P</i>	96	68.0	32	67.0
<i>I</i>	96	72.2	32	70.1

Note.—Abbreviated: *S* = standard, nonpersonal sentence stems, *P* = personal pronoun sentence stems, *I* = impersonal form—personal pronouns in the *P* sentence stems replaced with random proper female names.

TABLE 3
INTERCORRELATION MATRICES

Variable	<i>P</i>	<i>S</i>	<i>I</i>	<i>SD</i> Scale
Initial sample				
<i>P</i>	1.00	.40*	.46*	-.34*
<i>S</i>		1.00	.04	-.24*
<i>I</i>			1.00	.03
Cross-validation sample				
<i>P</i>	1.00	.41*	.24	No data
<i>S</i>		1.00	-.03	No data
<i>I</i>			1.00	No data

Note.—Abbreviated: *S* = standard, nonpersonal sentence stems, *P* = personal pronoun sentence stems, *I* = impersonal form—personal pronouns in the *P* sentence stems replaced with random proper female names.

* $p \leq .05$.

for the form scores, again, were significantly different in the ascending order *S*, *P*, *I* (see Table 2).

Correlations were obtained between the forms in both the first and second phases of the experiment (see Table 3). It can be seen that Form *P* correlated with Form *S* moderately, as reasonably might be expected of two selected halves of the same test. This was true in both the initial and the cross-validation samples. Form *P* and Form *I* also correlated moderately on the initial sample which was expected due to their similarity. This significant correlation was not replicated, although a finding of near significance was obtained. Form *S* did *not* correlate with Form *I* in either sample, which gave rise to the speculation that whatever was similar between Forms *P* and *I* was not that which was similar between Forms *P* and *S*.

Forms *P* and *S* also negatively correlated with the Marlowe-Crowne *SD* scale. This was expected, indicating that the higher was an *S*'s *SD* scale score, the lower was his standard (*P* or *S*) Rotter score. As predicted, a higher deference to social desirability (higher *SD* scale score) was related to less maladjusted (lower) ISB scores. However, it may be argued that when the necessity to express *personal* feelings was removed from the ISB (as was attempted in Form *I*) the responses were influenced correspondingly less by social desirability. Such was the case in this study.

Form I did not correlate significantly with the *SD* scale. Form I also elicited the highest, or most maladjusted, mean ISB scores of any of the three forms.

REFERENCES

- CROMWELL, R. L., & LUNDY, R. M. Productivity of clinical hypotheses on a sentence completion test. *Journal of Consulting Psychology*, 1954, 18, 421-424.
- ELLIS, A. A comparison of the use of direct and indirect phrasing in personality questionnaires. *Psychological Monographs*, 1947, 61(3, Whole No. 284).
- GETZELS, J. W., & WALSH, J. J. The method of paired direct and projective questionnaires in the study of attitude structure and socialization. *Psychological Monographs*, 1958, 72(1, Whole No. 454).
- MELTZOFF, J. The effect of mental set and item structure upon response to a projective test. *Journal of Abnormal and Social Psychology*, 1951, 46, 177-189.
- ROTTER, J. B., & RAFFERTY, J. E. *Manual: The Rotter Incomplete Sentences Blank. College form*. New York: Psychological Corporation, 1950.
- SACKS, J. M. Effect upon projective responses of stimuli referring to the subject and to others. *Journal of Consulting Psychology*, 1949, 13, 12-21.
- STEIN, M. I. The use of a sentence completion test for the diagnosis of personality. *Journal of Clinical Psychology*, 1947, 3, 47-56.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 24, 1967)

ASSOCIATIVE RESPONSE COMPETITION IN SCHIZOPHRENIA¹

ROBERT KAPCHE

California State College at Long Beach

The present study investigated Mednick's hypothesis that schizophrenics are especially susceptible to weak associative response competition. Groups of schizophrenics and college students (15 per group) each learned one of three paired-associate word lists. Weak intralist associative response competition was varied across the lists. Results showed a general performance inferiority for the schizophrenic groups. A significant list effect was also found with several of the learning measures. There was no support for the hypothesis, however. Normal and schizophrenic Ss were equally affected by the response competition.

Mednick (1958) has presented an hypothesis based on Hullian concepts as a partial explanation of the unintelligible verbal behavior of schizophrenics. He stated that schizophrenics have an intense anxiety level. The anxiety tends to raise remote-response tendencies to stimuli above the evocation threshold. In verbal behavior, this results in disruptions of established verbal habits due to associative response competition. Since these disruptions often reduce anxiety, the thinking and speaking of remote associations become habitual defenses against anxiety. The purpose of this study was to investigate this hypothesized relative susceptibility of schizophrenics to weak associative response competition, by means of a verbal learning task.

Previous experiments using verbal learning tasks to test Mednick's hypothesis have yielded inconsistent results. Both support (e.g., Kausler, Lair, & Matsumoto, 1964; Spence & Lair, 1965) and nonsupport (e.g., Donahoe, Curtin, & Lipton, 1961; Spence & Lair, 1964) have been reported. None of these studies, however, directly investigated the effects of weak associative response competition arising from preexperimental verbal habits. This is the type of competition stressed by Mednick. Since his hypothesis is that schizophrenics have intense anxiety, an

appropriate test of the hypothesis would be to compare schizophrenics and controls on a task on which it could be assumed that little associative response competition would be present under average levels of anxiety.

The present experiment studied the effects of weak intralist associative response competition on the learning of paired-associate word lists by schizophrenic and normal Ss. An interaction between S groups and the lists was predicted, with the schizophrenics' performance becoming relatively more inferior to the normals' as the level of associative competition increased over the lists. It was expected that the manipulated response competition would be so weak that it would not disrupt the performance of the normals.

METHOD

Materials

Three paired-associate word lists were devised, each containing five pairs. The stimuli were constant across the lists and were approximately equal in frequency, according to the Thorndike-Lorge (1944) norms.

In List 1, the control condition, there was minimal associative strength for all possible relationships between stimuli and responses. With the two remaining lists each response was a weak associate to one of the stimuli, but was not paired with it. In Lists 2 and 3, the associated response had been given an average of 11.00 and 19.6 times, respectively, according to Russell and Jenkins' (1954) norms.

The Russell and Jenkins' (1954) norms were combined with those of Bousfield, Cohen, Whitmarsh, and Kincaid (1961) and an additional study by the author in order to minimize as much as possible all other associative relationships among the items on the lists. The author's study collected associations to 11 stimuli from 110 college students in introductory psychology classes.

The sets of responses were also equated for word

¹ This study is based on a master's thesis submitted to Northwestern University. The author wishes to thank his thesis advisor, W. A. Hunt, and the staff at the Downey Veterans Administration Hospital for their cooperation and aid. During the collection of the data, the author was the recipient of a National Science Foundation Fellowship. The study was presented in part at Western Psychological Association meetings, May 5, 1967, San Francisco.

frequency, number of syllables, and number of words in each list beginning with the same letter. The lists are presented in Table 1.

Subjects

Forty-five schizophrenic *Ss* were drawn from the population of the Veterans Administration Hospital at Downey, Illinois. Forty were males. Most had histories of repeated hospitalizations and were diagnosed as chronic undifferentiated schizophrenics. The selection criteria were as follows: at least 8 yr. of education, no history of a psychiatric diagnosis other than schizophrenia, no indications of CNS pathology, no histories of alcoholism or drug addiction, and the willingness to cooperate and the ability to understand the instructions, as estimated by ward personnel. Most of the patients were lightly medicated. No systematic drug-level control was attempted, since Donahoe et al. (1961) found drug level to be unrelated to performance, and *E* found patients would often avoid taking their medication. A few potential *Ss* were not used, however, due to their being over-medicated. The mean age was 39.38 yr.; the mean educational level was 11.87 yr. The three subgroups of schizophrenics were equated for age and education.

The normal comparison group consisted of college sophomores fulfilling a course requirement. These *Ss* were assigned randomly to the conditions.

Comparing two such psychiatrically dissimilar groups probably should have accentuated any $S \times$ List interactions present. The two groups probably did differ on mean IQ. The available evidence indicates, however, that a mean difference between groups of about 40 points is needed before a clear difference in paired-associate learning is found, with the brighter group then showing superior performance (Zeaman & House, 1967). It is very doubtful that such a large difference could be found here, since the difference in mean educational level was not extreme. The older age of the schizophrenic group would also tend to retard performance (e.g., Canestrari, 1966), although Canestrari's study did not find response competition to be affected by age.

Procedure

Each *S* learned one of the lists. The *S* saw the stimulus alone for 2 sec., then the stimulus and

response together for 2 sec. Then the next stimulus was presented immediately. There was no intertrial interval, as an asterisk above the fifth response marked the end of each trial. Four orders of the pairs were used. Standard paired-associate instructions were given to the normal *Ss*. These instructions were lengthened and simplified for the schizophrenics, and efforts were made to relax them before the task was begun. The learning criterion was two consecutive correct anticipations of the list.

RESULTS

The means of the data to be discussed are presented in Table 2. The data were subjected to 2×3 factorial analyses of variance, the factors being *S* categories and degree of intralist associative competition across lists. The first analysis concerned trials to criterion. Due to great differences in variance between the normal and the schizophrenic groups on this measure ($F_{\max} = 14.28$, six groups, $df = 14$ per group, $p < .01$), a logarithmic transformation to the base 10 was used on this data. The subsequent analysis showed significant main effects for *S* categories ($F = 38.13$, $df = 1/84$, $p < .001$) and for lists ($F = 4.00$, $df = 2/84$, $p < .05$). The schizophrenics took more trials to criterion than did the normals. Scheffé's test for orthogonal comparisons (Edwards, 1962) revealed that the list effect was due to the poorer performance on Lists 2 and 3, relative to List 1. There was no support for an interaction between *S* categories and lists ($F < 1.00$).

Analyses of both the mean number of correct responses and the mean number of omissions per trial yielded the same general results. Again, there were significant effects for *S* categories and for lists and no noticeable trends toward interactions. The schizophrenics averaged fewer correct responses and more response omissions. Performance was worst on List 2, followed by List 3.

Analysis of the mean number of overt errors of all types per trial showed somewhat different results. Here the only significant effect was for *S* categories ($F = 15.12$, $df = 1/84$, $p < .01$), with the schizophrenics averaging more errors than the normals. A weak trend toward an interaction was also present. The trend was contrary to the predictions, though, as it showed the normal *Ss* tending to make more errors as the level of

TABLE 1
PAIRED-ASSOCIATE ITEMS

Stimulus	Response		
	List 1	List 2	List 3
1. SWIFT	Load (2) ^a	Stream (5)	Current (2)
2. BLOSSOM	Bread (1)	Pretty (3)	Smell (5)
3. STEM	Dark (3)	Lamp (2)	Branch (4)
4. CARPET	Boy (4)	House (1)	Walk (3)
5. PRIEST	Mountain (5)	Holy (4)	Black (1)

^a The number in parentheses corresponds to the number of the stimulus paired with it.

TABLE 2
MEANS OF PERFORMANCE MEASURES FOR SCHIZOPHRENICS AND NORMALS ON THE
PAIRED-ASSOCIATE LISTS

Measure	Group					
	Schizophrenic			Normal		
	List			List		
	1	2	3	1	2	3
Trials to criterion (raw scores)	18.07	25.80	20.13	6.53	10.60	8.13
Mean correct	3.24	2.69	2.80	4.06	3.41	3.69
Mean omissions	1.17	1.77	1.66	.75	1.27	.93
Mean errors	.62	.54	.53	.19	.30	.37
Mean percentage of response errors that were associative errors	—	.42	.31 ^a	—	.41 ^b	.53 ^b

^a $N = 14$.

^b $N = 11$. $N = 15$ in all other cells.

response competition increased from List 1 to List 3, while the schizophrenics did the opposite.

In order to test for possible effects of associative competition more directly, all errors consisting of the selection of one of the four incorrect responses were tabulated in Lists 2 and 3. The percentage of these errors that consisted of the selection of the associative response was also found. If associative competition were not present, the percentage of associative errors should not have been significantly greater than 25%. Results showed, instead, that 42% of all response errors were also associate errors, a difference significantly greater than 25% with a two-tailed t test ($p < .01$). A subsequent 2×2 analysis of variance on Ss' percentage of associative errors on Lists 2 and 3 failed to reveal any significant effects.

Response competition arising from extra-experimental sources seemed negligible. Only nine such responses were emitted by 90 Ss, seven of these by the normal Ss. No S emitted more than one.

DISCUSSION

It is apparent that the present study offers no support for Mednick's (1958) hypothesis that schizophrenics will be especially susceptible to weak associative response competition. None of the predicted interactions was significant, or even in the direction consistent with the hypothesis. The most direct

test of associative competition failed to differentiate the S categories. Perhaps the most striking feature of the data was the relative similarity in performance of the schizophrenics and the normals across the three lists.

These results also do not support the positions expressed in two recent reviews. Both Huff (1964) and Lang and Buss (1965) accepted the hypothesis that associative competition is stronger in schizophrenics than in normals.

The performance decrement of the schizophrenics found here was reflected in a lower rate of correct responses and higher rates for both response omissions and overt errors. These results differ from two recent studies. Donahoe et al. (1961) stated that the inferior performance of the mostly chronic schizophrenics they tested was due largely to a general failure to respond. Spence and Lair (1964), on the other hand, pointed out a tendency for their relatively acute sample of schizophrenics to make more overt errors, and to show performance equivalent to the normal controls when comparisons involving correct responses were made.

Noting this discrepancy, Spence and Lair suggested that length of current hospitalization may be related to performance differences among schizophrenics, with acute patients tending to make overt errors and chronic patients tending not to respond. This hypothesis was tested in the present study by correlating months of current hospitalization with both

the omission and the error rates. Pearson r 's were tabulated for each list separately, then averaged. However, neither correlation coefficient differed significantly from zero. Thus, Spence and Lair's hypothesis was not confirmed.

Despite the negative findings of Donahoe et al. (1961) concerning drug effects, medication may have reduced the anxiety of the schizophrenics enough to significantly lessen the strength of the associative competition. Indirect evidence against this hypothesis, however, is found in a study by Buss and Daniell (1967). They also reported results contrary to Mednick's hypothesis, as they found no differences in amount of stimulus generalization between groups of schizophrenics either on medication or on placebos compared with a group of normals. Their experimental design should now be used to study associative competition in schizophrenia.

REFERENCES

- BOUSFIELD, W. A., COHEN, B. H., WHITMARSH, G. A., & KINCAID, W. D. The Connecticut free association norms. Technical Report No. 35, 1961, University of Connecticut, Contract Nonr 631(00), Office of Naval Research.
- BUSS, A. H., & DANIELL, E. F. Stimulus generalization and schizophrenia. *Journal of Abnormal Psychology*, 1967, 72, 50-53.
- CANESTRARI, R. E. The effects of commonality on paired associate learning in two age groups. *Journal of Genetic Psychology*, 1966, 108, 3-7.
- DONAHOE, J. W., CURTIN, M. E., & LIPTON, L. Interference effects with schizophrenic subjects in the acquisition and retention of verbal materials. *Journal of Abnormal and Social Psychology*, 1961, 62, 553-558.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1962.
- HUFF, F. W. Learning and psychopathology. *Psychological Bulletin*, 1964, 61, 459-468.
- KAUSLER, D. H., LAIR, C. V., & MATSUMOTO, R. Interference transfer paradigms and the performance of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, 69, 584-587.
- LANG, P. S., & BUSS, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology*, 1965, 70, 77-106.
- MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, 55, 316-327.
- RUSSELL, W. A., & JENKINS, J. S. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, ONR Contract N8 ONR 66216.
- SPENCE, J. T., & LAIR, C. V. Associative interference in the verbal learning performance of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, 68, 204-209.
- SPENCE, J. T., & LAIR, C. V. Associative interference in the paired-associate learning of remitted and nonremitted schizophrenics. *Journal of Abnormal Psychology*, 1965, 70, 119-122.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.
- ZEAMAN, D., & HOUSE, B. J. The relation of IQ and learning. In R. M. Gagné (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967.

(Received August 7, 1967)

THE HYSTERIA AND PSYCHASTHENIA CONSTRUCTS AS AN ALTERNATIVE TO MANIFEST ANXIETY AND CONFLICT-FREE EGO FUNCTIONS¹

JAMES STEELE

San Jose State College

A total of 128 hysteric, psychasthenic, and control Ss, selected with the Hysteria and Psychasthenia scales of the MMPI, were employed to duplicate the performances of the "conflict-free" cognitive styles of leveling and sharpening and scanning under anxiety arousal. Psychasthenics performed like sharpeners and scanners on the Embedded Figures and Stroop Color-Word Tests, while hysterics performed like levelers. Duplication of the performance of "conflict-free" cognitive styles by anxiety-based personality defenses and the superior performance of the psychasthenic or "high-anxious" group on the complex Embedder Figures Test suggested modifications in the positions of Gardner on cognitive styles and Taylor on the Manifest Anxiety scale.

An investigation of the influence of hysteric and psychasthenic defensive styles upon the learning or performance of simple and complex tasks was undertaken in this study.

The purpose of the investigation was to demonstrate that hysteric and psychasthenic defenses can account for some of the findings of J. Taylor and K. W. Spence (1952) in relation to the task performances of Ss high and low on the Taylor Manifest Anxiety (*MA*) scale (J. A. Taylor, 1953) and the findings of Gardner, Holzman, Klein, Linton, and Spence (1959) in relation to the task performances of Ss having the conflict-free cognitive styles of leveling, sharpening, and scanning.

It is a basic hypothesis of this study that levelers, whom Klein (1954) described as having hysterical defenses and tendencies to repress, will be primarily low scorers on the Taylor *MA* scale, while sharpeners and scanners, who have obsessional defenses, will be high scorers (Gardner et al., 1959). This hypothesis would be based in part on low nonsignificant or negative correlations between the Hysteria and Psychasthenia scales of the MMPI and the reported correlations of .80 or .90 between the Psychasthenia scale and the Taylor *MA* scale (Erickson & Davids, 1955).

¹This paper is an abridgment of a dissertation submitted in partial fulfillment of the PhD degree and sponsored by Irwin Sarason at the University of Washington, 1961.

Gardner et al. (1959) reported that levelers who could be classified as low on the Taylor *MA* scale perform poorly on the Witkin Embedded Figures Test because of their inability to articulate stimulus configurations. They also perform poorly on tasks requiring memory for detail and size judgments. Sharpeners, who are primarily professional women, perform in a superior fashion on the Embedded Figures Test, and scanners, who are primarily male Ss, perform in a superior fashion on tasks requiring size judgments. Scanners have great difficulty, however, with a simple task, that is the Stroop Color-Word Test (Stroop, 1935). There is some evidence that levelers have much less difficulty with this task (Gardner et al., 1959).

Gardner et al.'s (1959) findings, then, suggest that sharpener and scanner Ss, who are potentially high on the Taylor *MA* scale, can perform in a superior fashion on complex tasks such as the Embedded Figures Test, while low-anxious Ss (levelers) do poorly on a complex task.

This finding is, by inference, a direct contradiction of the original Taylor and Spence position where high-anxious Ss reportedly have inferior performance on complex tasks as a function, in Hullian terms, of the high drive state and activation of task-irrelevant and competing habits (J. Taylor & K. W. Spence, 1952). More recently, as a result of inconsistencies in her findings, Taylor has

stated that her theory is incomplete with regard to properties other than drive level which may differentiate groups with extremely high scores on the *MA* scale from those with extremely low scores (J. Spence, 1963).

A clarification of both the Taylor and Gardner-Klein positions is suggested by the original interpretation of the Psychasthenia scale of the MMPI as a measure of obsessional traits as well as manifest anxiety (Hathaway & McKinley, 1951).

An intellectual feature of the obsessive, in addition to manifest anxiety, is vigilant and compulsive attention to detail, retentiveness, and what may be termed overinclusion of stimuli, while hysterics are contrasted as excluders who avoid or forget stimuli (Cameron & Margaret, 1951).

These theoretical and clinical positions have received experimental support from the studies of Erickson and Davids (1955), Truax (1957), Truax and Martin (1957), Ruebush (1960), Gordon (1957), and Barron (1953) where, in general, *Ss* classified as hysterics or repressors tended to forget failures and were less detailed in their memory and perceptual functions, while the psychasthenic or sensitized *Ss* tended to remember unpleasant associations and displayed more differentiated memory and perceptual functions.

The general exclusion capacity of the hysteric was suggested also in a recent study by Wollitsky (1967) in which he found that *Ss* who obtained low-interference scores on the Stroop Color-Word Test denied a relationship between cigarette smoking and lung cancer to a significant degree when compared to the high-interference group. He concluded that the facility for ignoring intrusive words extends to the area of belief.

Messick and Fritzky (1963), commenting on studies of conceptual styles by Witkin, Dyk, Fatterson, Goodenough, and Karp (1962), and Kagan, Moss, and Sigel (1963), noted that these studies have produced a dimension of the analytic in contrast to the global attitude. Witkin (1965) related global and analytic conceptual styles to the defenses of repression and isolation, respectively. Rorschach interpretation has long used global or whole responses as indicative of hysterical traits and detail responses as indicative of obsessional traits (Schafer, 1948).

From the studies just mentioned one could infer that more exact predictions of task performance could be made if attention were paid both to defensive style and to task requirements.

To test this inference it was predicted that psychasthenic *Ss* or high-anxious *Ss* would perform in a superior fashion on a complex task, the Witkin Embedded Figures Test under stress, in contrast to a Taylor prediction of inferior performance. It was also predicted that hysteric *Ss*, who are manifestly low-anxious *Ss*, would perform in an inferior fashion on the same complex task and in a superior fashion a simple task, the Stroop Color-Word Test, in contrast to the Taylor position. Superior or inferior performances would thus depend upon whether the task is compatible with the intellectual differentiation, retention, and attention to detail of the psychasthenic, as in the case of the Witkin test, or compatible with the lack of retentiveness, differentiation, and attention to detail of the hysteric, as in the case of the Stroop test.

The introduction of stressful instructions to increase levels of anxiety was included to increase the defense and magnify these predicted results. Statistically, this would occur as a significant Defense \times Instructions interaction.

Duplication of the performance of levelers and of sharpeners and scanners by hysterics and psychasthenics could be predicted by Gardner et al. (1959), but these authors would not predict systematic changes under experimental stress, as cognitive styles are conflict free and not derived from anxiety reduction.

In testing the hypothesis above, groups of hysteric, psychasthenic, and control *Ss* were exposed to the Embedded Figures and Stroop Color-Word tests. It was predicted that control *Ss* would not show systematic changes predicted for the experimental groups.

Controls for intelligence were included to prevent the confounding of intelligence and defenses.

METHOD

Selection of Subjects

The variables of concern to the study were incorporated into a $4 \times 2 \times 2 \times 2$ factorial analysis of

variance design in which defenses were varied in 4 ways, sex in 2, instructions in 2, and order of task administration in 2 ways. It can be seen from this arrangement that 32 cells are involved in the design and that, with a minimum of 4 Ss to a cell, 128 Ss would be required for the design. The design required an analysis of variance for scores obtained on both the Witkin and Stroop tests.

The Ss were selected from 942 beginning students in psychology and sociology on the basis of a score in the upper or lower 30% of a distribution of scores obtained from the administration of the *Hy* and *Pt* scales of the MMPI to these students. The Ss were administered the *Hy* and *Pt* scales by a representative of *E*. Two weeks later the same students were asked to submit their names for a one-in-five random drawing of Ss to standardize some tests.

Four combinations of defenses were obtained from the *Hy* and *Pt* measurements, individuals who were high on one scale and low on the other or individuals high on both scales or low on both scales. High on the *Hy* scale was a raw score of 24 or above and 19 or above on the *Pt* scale. Lows on both scales were 15 and below and 9 and below, respectively. It was difficult to obtain Ss who were high and low on both scales.

Independent Variables

The major independent variable was neutral versus stressful instructions. Other independent variables were defenses, order, and sex. Neutral or stressful instructions were given prior to the standard instructions which are provided for the Witkin Embedded Figures or Stroop Color-Word Tests. One major change in the standard instructions for the Witkin test was made. The Ss were not informed that they could ask to look at the simpler designs again if they so desired. This change was introduced on the basis of Ruebush's (1960) finding that high-anxious Ss asked to see the designs more frequently than low-anxious Ss. In addition, Gardner et al.'s (1959) data indicated that levelers forgot the design more frequently than sharpeners. To eliminate the possible effects of these tendencies on performance, Ss were reexposed to a simple design every 60 sec. until they reached a solution. Failure-induced anxiety was avoided by informing Ss that the usual procedure was to let them see the design again at regular intervals.

Dependent Variables

Dependent variables were scores obtained from performance on the Witkin Embedded Figures Test and the Stroop Color-Word Test.

Scores for the analysis of the Witkin test were derived by obtaining the mean solution time per figure among the 24 figures for each S.

Scores for the Stroop Test were obtained by the method suggested in the Gardner et al. (1959) monograph. Each S's score was derived by determining the difference between his predicted and obtained scores on the third phase of the Stroop test. This score is a measure of interference. Predictions were

made on the basis of a regression equation derived from the correlation between the second and third phases of the Stroop test. Scores obtained directly by time on the third phase of the test were also employed.

Each S was administered both of the task discussed above to make the most efficient use of Ss. Analysis of the data for a possible order effect was arranged by alternating the order of administration of the tasks.

Sex differences on the Witkin test have been established by Witkin (1950). Control for this anticipated difference in performance was obtained by assigning an equal number of each sex to each experimental condition.

Intelligence

Control for intelligence was accomplished by restricting the range of intelligence of Ss markedly below the mean obtained American Council on Education Test of Intelligence (ACE) scores for experimental and control groups used in the study (revised by the University of Washington Counseling Center, 1960). This resulted for the most part in the removal of several male and female hysterics. After selection or exclusion on the basis of these scores, *t* tests revealed no significant differences among mean total, quantitative, and verbal ACE scores of Ss assigned to the various experimental and control groups.

Experimental Arrangements

The Ss were administered the Witkin and Stroop tests individually. Each S was seated facing a 2 × 3 ft. vertical plasterboard placed on a table upon which were mounted two identical blue notebooks containing the respective tests. The *E* timed S's performance with a stopwatch and recorded his score on record sheets prepared for that purpose.

The Ss were screened for color blindness before they received instructions. After this inquiry, Ss received either the natural or stressful instructions followed by the standard instructions for the test.

Stressful instructions were as follows:

Each of these notebooks that you see in front of you contains a test that has been used to evaluate effective decision-making in a short period of time. Originally the tests were used to select executive personnel. As you know, a person in an executive position often has to make an effective decision without having a lot of time to collect information on a subject. The tests have also been used in test batteries designed to measure psychological adjustment since effective decision-making is one measure of psychological adjustment. Our present interest, however, is to use the tests to predict a student's grade point average during his senior year. Before we begin, could you tell me the GPA that you expect to earn during your senior year.

(The prediction was obtained, and then the student was asked his current GPA.) Following this

exchange, S was asked if he planned to graduate from this university.

Neutral instructions were as follows:

The two notebooks you see in front of you each contain a test that has been developed for general experimental purposes. At this point we are simply interested in establishing some norms or averages for these tests since we don't know how college students will perform on them.

The effectiveness of the stressful instructions was evidenced by very noticeable increases in random motor behavior, for example, shifting in the chair, arm, hand, and facial movements.

Administration of the tests under neutral conditions should be interpreted as a moderately stressful condition since the tests are timed and relatively difficult to perform.

RESULTS

An analysis of variance conducted with the four defense categories resulted in a significant *F* as a result of the superior performances of male Ss on the Witkin test and a significant Instructions \times Sex interaction (table not included). The superior performance of male Ss was consistent with Witkin's (1950) findings. The significant Instructions \times Sex interaction can be attributed to the superior performance of men under neutral instructions, while females tended to display superior performance under stressful instructions.

Since hysteric and psychasthenic performance differences were obscured by the less extreme performances of the control groups in the analysis of the four groups (see Table 1), it is necessary to turn to the analysis of the high-low (hysteric) and low-high (psychasthenic) groups alone to provide support for the basic hypotheses advanced at the beginning of this study.

TABLE 1

MEAN SCORES: TOTAL EMBEDDED FIGURES FOR EIGHT EXPERIMENTAL GROUPS

	Hysteric				Psychasthenic			
	Control Ss				Control Ss			
	LLN	LLS	HLN	HLS	LHN	LHS	HHN	HHS
<i>M</i>	33.93	29.62	29.81	40.93	33.68	25.	33.06	35.24

Note.—Abbreviated: L = low, H = high, N = neutral instructions, S = stressful instructions.

TABLE 2

ANALYSIS OF VARIANCE FROM SOLUTION TIME OF 24 EMBEDDED FIGURES FOR TWO DEFENSE MODES: HYSTERIA AND PSYCHASTHENIA ONLY

Source of variance	<i>df</i>	<i>MS</i>	<i>F</i>
Between subjects			
Defense (A)	1	581.	3.489
Instructions (B)	1	23.76	—
Sex (C)	1	669.51	4.021*
Order (D)	1	301.89	1.81
A \times B	1	1,571.	9.435**
A \times C	1	142.	.852
A \times D	1	180.	1.081
C \times D	1	722.53	4.339
B \times C	1	1,130.73	6.791**
B \times D	1	425.55	2.555
A \times B \times C	1	46.65	.280
A \times B \times D	1	17.04	.102
B \times C \times D	1	628.67	3.775
A \times C \times D	1	76.40	.458
A \times B \times C \times D	1	258.27	1.551
Within subjects	48	166.5	—
Total	63		

* $p < .05$.

** $p < .01$.

Basic interest should center upon the highly significant Defense \times Instructions interaction (Table 2) which is significant beyond the .01 point of probability after correction for heterogeneity of variance (Lindquist, 1953). This significant interaction provides major support for two basic hypotheses of the study: that the personality defenses of hysteria and psychasthenia can be used to predict performance on the Embedded Figures Test and that one manifestation of this relationship is a significant Defense \times Instructions interaction.

Mean scores for the four defensive modes when performing under neutral and stressful conditions are presented in Table 1. Inspection of this table indicates that the significant interaction obtained from the analysis of the hysteric and psychasthenic groups can be attributed to the poorer performance of the hysteric group under stress and the improved performance of the psychasthenic group under the same condition. In this case, the interaction can be attributed to the inferior performance of the females under neutral conditions. In addition, the hysteric males are inferior under stressful conditions.

Differences in performance which are characteristic of the hysteric and psychasthenic

TABLE 3

MEAN OBTAINED SCORES IN SECONDS OF EIGHT EXPERIMENTAL GROUPS ON STROOP COLOR-WORD TEST

	Control		Hysteric		Psychasthenic		Control	
	LLN	LLS	HLN	HLS	LHN	LHS	HHN	HHH
M	99.06	99.62	94.25	96.87	107.43	100.7	95.62	104.50

Note.—Abbreviated: L = low, H = high, N = Neutral instructions, S = stressful instructions.

groups under neutral and stressful conditions do not occur in the low hysteric-psychasthenic or control group (see Table 1). This finding adds additional significance to the systematic fluctuations of the hysteric and psychasthenic groups. It is of interest, also, to note that the high hysteric-psychasthenic control group does not differ under neutral and stressful conditions in the same systematic manner as the two experimental groups.

As an additional point it is obvious from the preceding data that the predicted differences between the hysteric and psychasthenic groups did not materialize until the introduction of stress. While this finding fails to confirm one hypothesis of the study—that hysterics and psychasthenics would differ under neutral conditions—it may be a function of controls exercised for intelligence in this study.

Stroop Color-Word Test Results

Inspection of the distribution of scores on the Stroop Color-Word Test indicated heterogeneity of variance. The procedure suggested by Lindquist (1953) was also applied to the data obtained from this analysis.

Although the original design of this study provided only for an analysis of scores based upon the difference between the predicted and obtained scores on the third phase of the Stroop test as suggested by Gardner et al. (1959), an additional analysis was conducted to compensate for possible shortcomings in the scoring system listed above. The additional analysis used time scores for the third phase of the Stroop test as a direct measure of interference.

The only significant finding resulting from an analysis of variance of all four experimental groups using the first method, differ-

ence scores (table not included), was the order effect. The significant order effect was a result of the higher interference scores obtained under Order I administration of the tests. This order involved the administration of the Stroop test as the second test and might be interpreted as the effect of fatigue or accumulated anxiety over performance on the preceding test. Inspection of the performance of hysterics and psychasthenics revealed the mean difference scores of the hysteric and psychasthenic groups differ in the predicted manner, but these differences failed to reach statistical significance (table not included).

Table 3 presents the mean time rather than difference scores of the entire eight groups of Ss of interest to the study. Table 4 presents the results of an analysis of variance using scores from the hysteric and psychasthenic group only. This analysis indicates that the psychasthenic group displayed higher interference.

The significant main effect between defenses is not accompanied by a significant Defense \times Instructions interaction (see Table 4). None of the other variables shows significant effects. Support for one basic hypothesis of the study, that the psychasthenic group would show higher interference and the hys-

TABLE 4

ANALYSIS OF VARIANCE OF STROOP COLOR-WORD TEST OBTAINED SCORES: HYSTERIA AND PSYCHASTHENIA GROUPS ONLY

Source of variance	df	MS	F
Between subjects			
Defense (A)	1	1,164.51	5.34*
Instructions (B)	1	66.01	—
Sex (C)	1	135.89	—
Order (D)	1	708.89	3.25
A \times B	1	348.50	1.59
A \times C	1	461.85	2.118
A \times D	1	16.60	—
C \times D	1	437.60	2.00
B \times C	1	94.30	—
B \times D	1	321.3	1.47
A \times B \times C	1	74.10	—
A \times B \times D	1	212.12	.973
B \times C \times D	1	98.37	—
A \times C \times D	1	7.15	—
A \times B \times C \times D	1	166.56	—
Within subjects	48	218.06	
Total	63		

* $p < .05$.

teric, a lower interference on the third phase of the interference test, was obtained.

Of interest, also, are the relationships between the means where the lowest means (lowest interference) are obtained by Ss in the high *Hy* groups, who are in effect high scorers on the *Hy* or repression variable.

DISCUSSION

The superior performances of the psychasthenic group on the Embedded Figures Test under stress and the superior performance of the hysteric group on the Color-Word Test support the importance attached to personality defenses at the beginning of this study.

The significance of the systematic differences is also supported by the control groups, who did not show superior or inferior performance from task to task.

Statistically the predicted Defense \times Instructions interaction has been confirmed. Performance of the hysteric group showed a significant decrement under stress, while the psychasthenic group tended to show a significant increment in performance under stress on a complex task, the Embedded Figures Test.

In terms of defensive processes, the superior capacity of the hysteric to exclude stimuli and reduce interference from competing stimuli appears to be an asset in the Color-Word Test and a handicap in the Embedded Figures Test, while the inclusive capacity of the psychasthenic produces directly opposite results when the inclusive defenses are activated by stress.

As a general conclusion it appears that there has been a tendency toward oversimplification of the effects of personality dimensions upon task performance. It is difficult to assess these effects by using single variables such as drive level as originally suggested by J. Taylor and K. W. Spence (1952). Specific consideration of task performance as an interaction between personality defenses and task requirements has made it possible to produce superior performance on a complex task by high-anxious (psychasthenic) Ss. Since these findings directly contradict the Taylor-Spence position, this study supports interpretation of high and low scores on the Taylor scale as measures of defensive

processes rather than of drive. Insofar as tests of Hullian theory are tied to the Taylor scale, this study also indicates a need for modification of Hullian drive theory to take account of traditional clinically derived personality defenses.

An essential result of the study has been the demonstration of differences in performance as a function of personality defenses on tests employed by Gardner et al. (1959) to demonstrate cognitive controls. While the prediction that nonstress differences could be demonstrated on the Witkin and Stroop tests was not established for the Witkin test, the control for intelligence in this study and the consequent elimination of some low-scoring hysterics who might perform poorly on the Witkin test may be sufficient to explain the failure to duplicate Gardner et al.'s (1959) cognitive controls of leveling and sharpening and scanning under nonstress conditions.

A functional relationship has been obtained between the arousal of anxiety and performance in the hysteric and psychasthenic groups comparable to Gardner et al.'s (1959) levelers and sharpeners, respectively, on the Embedded Figures Test. This anxiety-based relationship casts doubt on the independence of cognitive controls from conflict and anxiety and suggests that cognitive controls may be a manifestation of the specific performance characteristics of a particular defensive style.

REFERENCES

- BARRON, F. Complexity as a personality dimension. *Journal of Abnormal and Social Psychology*, 1953, 48, 163-172.
- BERKOWITZ, L. Leveling tendencies and the complexity-simplicity dimension. *Journal of Personality*, 1957, 26, 743-746.
- CAMERON, M., & MARGARET, A. *Behavior pathology*. Boston: Houghton Mifflin, 1951.
- ERICKSON, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the Hysteric and Psychasthenic scales from the MMPI. *Journal of Abnormal and Social Psychology*, 1955, 50, 135-137.
- GARDNER, R., HOLZMAN, P., KLEIN, G., LINTON, H., & SPENCE, P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1, 4.
- GORDON, J. Interpersonal predictions of repressors and sensitizers. *Journal of Personality*, 1957, 25, 686-698.

- HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the MMPI*. New York: Psychological Corporation, 1951.
- KAGAN, J., MOSS, H. A., & SIGEL, I. E. Psychological significance of styles of conceptualization. In J. C. Wright & J. Kagan (Eds.), *Basic cognitive processes in children. Monographs of the Society for Research in Child Development*, 1963, 28(2, Whole No. 861), 73-112.
- KLEIN, G. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1954*. Lincoln: University of Nebraska Press, 1954.
- LINDQUIST, M. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MESSICK, S., & FRITZKY, F. J. Dimensions of analytic attitude in cognition and personality. *Journal of Personality*, 1963, 1, 347-369.
- RUEBUSH, B. R. Interfering and facilitating effects of test anxiety. *Journal of Abnormal and Social Psychology*, 1960, 60, 205-211.
- SCHAFER, R. *The clinical application of psychological tests*. New York: International Universities Press, 1948.
- SPENCE, J. TAYLOR. Learning theory and personality. In J. M. Wepman & R. W. Heine (Eds.), *Concepts of personality*. Chicago: Aldine, 1963.
- STROOP, J. R. Studies in interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, 18, 643-661.
- TAYLOR, J., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *Journal of Experimental Psychology*, 1952, 44, 6-64.
- TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, 48, 285-290.
- TRUAX, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *Journal of Abnormal and Social Psychology*, 1957, 55, 188-193.
- TRUAX, C. B., & MARTIN, B. The immediate and delayed effect of failure as a function of task complexity and personalization of failure. *Journal of Abnormal and Social Psychology*, 1957, 55, 16-20.
- WITKIN, H. A. Individual differences in ease of perception of embedded figures. *Journal of Personality*, 1950, 19, 1-15.
- WITKIN, H. A. Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology*, 1965, 70, 317-336.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- WOLITZKY, D. L. Cognitive control and cognitive dissonance. *Journal of Personality and Social Psychology*, 1967, 5, 486-489.

(Received August 3, 1967)

A PROCEDURAL CRITIQUE OF "DESENSITIZATION AND THE EXPERIMENTAL REDUCTION OF THREAT"

GERALD C. DAVISON

State University of New York at Stony Brook

It is suggested that conclusions from an experiment by Folkins, Lawson, Opton, and Lazarus be tempered with respect to two procedural flaws: (a) Their control for relaxation unpaired with aversive stimuli is unsatisfactory because *Ss relaxed while* viewing the test stress film; and (b) *Ss* in the "simulated desensitization" condition were not allowed to escape aversive images when they became disturbing, this error in design violating the principle in desensitization that aversive scenes be paired with relaxation, not with emotional upset. The results, therefore, cannot be taken as evidence either for or against the importance of any desensitization variables or combinations thereof.

The experiment by Folkins, Lawson, Opton, and Lazarus (1968) raises important questions regarding the largely neglected role of cognitive processes in "behavior therapy techniques" such as Wolpe's (1958) systematic desensitization procedure. Close examination of their methods, however, reveals two problems which seem to justify considerable caution in interpreting their experimental results.

Their principal surprising finding was that, on several measures, those *Ss* who had undergone an analogue of the clinical desensitization procedure did not fare as well as either those who were merely relaxed or those who rehearsed forthcoming stressful incidents without benefit of relaxation. Thus, a question Folkins et al. are asking is very close to that investigated by this writer (Davison, 1968); namely, is it necessary to pair anxiety-competing relaxation with imaginal aversive stimuli in order to effect reduction of fear (or stress), or can one account for the demonstrated efficacy of Wolpe's procedure in terms of relaxation alone and/or in terms of imagination of aversive scenes without relaxation ("extinction effects," or, as Folkins et al. termed it, "cognitive rehearsal")? The present writer found (as did Rachman, 1965, in a similar pilot study) that only when imaginal aversive stimuli are paired with relaxation, that is, during systematic desensitization, does significant reduction in avoidance behavior occur. The *Ss* who relaxed while imagining scenes *unrelated* to their fears and *Ss* who imagined aversive

scenes without relaxation did not improve significantly.

Although the Folkins et al. study is not intended as a replication of Davison (1968), the fact that their closely related findings are so radically different seems quite important. Let us therefore examine the procedures more closely.

In their Condition II—"Relaxation"—they hoped to evaluate the effects of relaxation alone (i.e., without pairing with aversive stimuli) in reducing stress reactions to an upsetting film. The *Ss* were given two sessions of relaxation training by means of a recording by A. A. Lazarus (upon which this writer based the training procedures in his own study), and in the third test session they were shown the film immediately following a third practice session with the relaxation tape. The film itself was preceded by these instructions:

It is felt that the relaxation you are now experiencing and the techniques you have learned to control these sensations can be used to your advantage to handle the somewhat stressful movie film you are about to see. Just remain as relaxed as you can. The more you are able to relax, the more you will enjoy the experience [p. 103].

Thus, one finds, *Ss instructed to relax during* the presentation of actual aversive stimulus material. This procedure is virtually identical to what has been occasionally reported (e.g., Clark, 1963) with clients who are either unable to generate clear images for systematic desensitization or with *Ss* undergoing what is usually referred to as "in

vivo desensitization" (cf. Davison, 1965). The crucial fact is that, in contrast to the "Relaxation Group" of Davison (1968), which imagined *unrelated* scenes while relaxed, Folkins et al.'s "relaxation group" seems to have experienced the apparently crucial *pairing* of anxiety-competing relaxation with stressful stimuli. It is submitted, therefore, that this is an unsatisfactory control for relaxation alone.

A second problem, which Folkins et al. alluded to, is that, because the "simulated desensitization" was conducted via tape recording, "Ss in the desensitization condition may have been sensitized rather than desensitized to the threat scenes [p. 111]." That is, unless the person being desensitized can signal to the therapist-*E* that a given scene is disrupting the relaxation, one is setting the stage for a pairing of imaginal aversive stimuli with emotional *upset*, rather than with emotional *calm*. Since the design of their Desensitization condition does not fulfill even the most basic requirements of Wolpe's (1958) original procedure, it does not seem reasonable to regard the outcomes of this particular experimental group as reflecting anything at all upon the efficacy of the technique or upon the variables underlying its efficacy. (See Davison, 1968, for a discussion of a similar problem in one of his control groups; also see Schubot, 1966, for a procedural modification which handles the problem.) Lang (in press) has independently voiced this concern.

In spite of what the present writer feels to be serious methodological flaws in this study, the questions raised about the role of cognitive processes in behavior therapy, and particularly in systematic desensitization, are intriguing and important. While this writer (Davison, 1968) interprets his main findings as support for a counterconditioning process underlying the beneficial outcomes of systematic desensitization, a later study, by Valins and Ray (1967), has raised a rather compelling alternative explanation for the beneficial outcomes of Wolpe's procedure. Valins and Ray suggested that, by relaxing in the face of hitherto disturbing images,

Ss reevaluate the significance of these events (i.e., change their cognitions) and subsequently (consequently as well?) manifest significant reduction in avoidance behavior. This is not dissimilar to London's (1964, pp. 130-131) Tolmanian interpretation of desensitization, whereby,

The closer the imagery comes to representing "real" experience of the most complete sort without being followed by the actual experience it simulates, the more the patient's expectation of disastrous action, with its disastrous consequences, is reduced. . . .

It is the task of future carefully designed experiments to ferret out the active ingredients of this fascinating procedure in such a fashion as to allow some choice among alternate explanations.

REFERENCES

- CLARK, D. F. The treatment of monosymptomatic phobia by systematic desensitization. *Behaviour Research and Therapy*, 1963, 1, 63-66.
- DAVISON, G. C. Relative contributions of differential relaxation and graded exposure to in vivo desensitization of a neurotic fear. In, *Proceedings of the 73rd annual convention of the American Psychological Association* 1965. Washington, D. C.: American Psychological Association, 1965.
- DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, 73, 91-99.
- FOLKINS, C. H., LAWSON, K. D., OPTON, E. M., JR., & LAZARUS, R. S. Desensitization and the experimental reduction of threat. *Journal of Abnormal Psychology*, 1968, 73, 100-113.
- LANG, P. J. The mechanics of desensitization and the laboratory study of human fear. In C. M. Franks (Ed.), *Behavioral therapies: Assessment and appraisal*. New York: McGraw-Hill, in press.
- LONDON, P. *The modes and morals of psychotherapy*. New York: Holt, 1964.
- RACHMAN, S. Studies in desensitization: I. The separate effects of relaxation and desensitization. *Behaviour Research and Therapy*, 1965, 3, 245-252.
- SCHUBOT, E. The influence of hypnotic and muscular relaxation in systematic desensitization of phobias. Unpublished doctoral dissertation, Stanford University, 1966.
- VALINS, S., & RAY, A. A. Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology*, 1967, 7, 345-350.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.

(Received August 3, 1967)

A REPLY TO DAVISON'S CRITIQUE

CARLYLE H. FOLKINS, KAREN LAWSON EVANS, EDWARD M. OPTON, JR.

AND RICHARD S. LAZARUS

University of California, Berkeley

We disagree with both of Davison's (1968) criticisms of our paper, "Desensitization and the Experimental Reduction of Threat" (Folkins, Lawson, Opton, & Lazarus, 1968).

Davison's first criticism is that our "relaxation" treatment included elements of desensitization. He writes, "The crucial fact is that . . . Folkins et al.'s 'relaxation group' seems to have experienced the apparently crucial *pairing* of anxiety-competing relaxation with stressful stimuli [Davison, 1968, p. 87]." This "procedural flaw" might therefore account for the fact that our relaxation treatment was as effective as our desensitization treatment, whereas Davison's (1965) Ss, in a design free from this "flaw," benefited only from his desensitization treatment, and not at all from his relaxation treatment.

In fact, however, it was Davison's procedure that permitted ". . . the apparently crucial *pairing* of anxiety-competing relaxation with stressful stimuli" during training. Davison (1965) exposed his Ss to a snake as a test for initial strength of snake avoidance. Then he administered experimental treatments: relaxation, desensitization, etc., and finally he readministered the snake-avoidance test. Davison's Ss knew, at the time of their relaxation treatment, that the purpose of the treatment was to reduce their fear of approaching snakes. Indeed, "The incentive for participation was [in part, desire] to be rid of a particular fear . . . [Davison, 1965, p. 18]." In consequence, Davison's Ss had ample opportunity and motivation to think about snakes during the relaxation training and hence to contaminate the relaxation treatment with elements of desensitization. In contrast, our design avoided this particular procedural flaw. Our Ss knew only that they would see a somewhat stressful film, but they knew nothing about the film content until they actually saw it. Thus, our procedure more adequately isolated relaxation from desensitization.

Our Ss were explicitly encouraged to make use of their relaxation training in dealing with the stressful film; Davison's Ss were implicitly encouraged to do so. In this respect the relaxation treatments of both Davison and the present writers resembled the clinical application of relaxation therapy, and both were clearly quite different from systematic desensitization therapy. The question remains open as to why our relatively "pure" relaxation treatment was effective, while Davison's relaxation treatment, which may have been contaminated to some extent by what he calls "in vivo desensitization," was ineffective. Our data remain a challenge to Rachman's (1967) conclusion, based on a review of empirical studies (including Davison's), that *pairing of aversive imagery* with relaxation during training is necessary in order to produce salutary effects.

Davison's second criticism was anticipated and dealt with in our original discussion where we pointed out that, since the procedure was standardized, Ss in the simulated desensitization condition could have experienced increased anxiety rather than relaxation while imagining the threat scenes. But this aspect of the simulated desensitization treatment is equally true of the cognitive rehearsal condition, in which the same threat scenes were presented without relaxation. Yet the beneficial effects of the cognitive rehearsal condition were greater than those of either our desensitization or relaxation conditions. Thus, our data are inconsistent with the possibility suggested by Davison that lack of procedural flexibility might have accounted for the failure of the full desensitization procedure to be superior to its relaxation or cognitive rehearsal components.

If Davison's criticisms were valid, they would vitiate the challenge which our data present to the reciprocal inhibition interpretation of the efficacy of desensitization therapy.

Since the data do remain an embarrassment to the reciprocal inhibition theory, other theoretical formulations must be explored. Our conclusion that the role of *cognitive processes* is probably an important one is a conclusion with which Davison seems to agree.

REFERENCES

- DAVISON, G. C. The influence of systematic desensitization, relaxation, and graded exposure to imaginal aversive stimuli on the modification of phobic behavior. Unpublished doctoral dissertation, Stanford University, 1965.
- DAVISON, G. C. A procedural critique of "Desensitization and the experimental reduction of threat." *Journal of Abnormal Psychology*, 1969, 74, 86-87.
- FOLKINS, C. H., LAWSON, K. D., OPTON, E. M., JR., & LAZARUS, R. S. Desensitization and the experimental reduction of threat. *Journal of Abnormal Psychology*, 1968, 73, 100-113.
- RACHMAN, S. Systematic desensitization. *Psychological Bulletin*, 1967, 67, 93-103.

(Received August 3, 1967)

STAFF ATTITUDES TOWARD PATIENT CARE AND TREATMENT-DISPOSITION BEHAVIOR

ALAN S. DEWOLFE, ROBERT P. BARRELL, AND FRED E. SPANER¹

Veterans Administration Hospital, Downey, Illinois

Treatment milieus (staff attitudes) were related to treatment-disposition behavior in four autonomous units with comparable staff and patients. Only measures on which units showed significant differences or ratios of these measures were analyzed. Attitudes toward rules ($p < .01$), informing patients ($p < .001$), impact of staff attitudes ($p < .01$), and patients' characteristics ($p < .05$), and measures of use of restraint and seclusion ($p < .01$), trial visits ($p < .05$), and maximum hospital benefit discharges ($p < .001$) met this criterion. The main focus was on three related attitudes interpreted as indicating openness, frankness, and trust in interactions with others and their significant relationships to an index of trust in patients' capabilities ($p < .05$ in all cases). These attitudes appeared sufficiently important in the treatment climate to merit further exploration.

Recently, considerable research energy has been turned to a multidimensional approach to measuring staff attitudes toward patient care in the psychiatric setting. The custodial versus treatment orientation has been subdivided into a number of more specific dimensions. With varying success, attempts have been made to relate staff attitudes to effectiveness in working in the psychiatric setting, for example, Carstairs, Gilbert, Heron, Levinson, and Pine (1957), Canter (1963), Cohen and Struening (1964), and Ellsworth (1965).

The present study evaluated the relationships between different treatment milieus, as measured by staff attitudes toward patient care, and treatment-disposition behavior in a large psychiatric hospital. Four autonomous treatment units, comparable in composition of staff and patients, provided the variation in treatment milieu or unit-staff attitudes toward patient care.

METHOD

Subjects

The attitudes of 418 staff members in the four units were evaluated three times over the period of 1 yr. Monthly measures for a 2-yr. period of treatment and disposition variables were analyzed.

¹ The authors wish to express appreciation to Winfred F. Hill of Northwestern University, who served as consultant, for his many helpful suggestions, and Albert Erlebacher of Northwestern University, for advice on statistical procedures.

These measures involved approximately 1,600 psychiatric patients under treatment in the four units.

Measures

Staff attitudes toward patient care were measured by the Philosophy of Treatment Form (POT), explained in Barrell, DeWolfe, and Cummings (1965). The units were compared on the following patient-care attitude areas: (A) interprets rules in an authoritarian manner, (B) thinks patients should be told their condition, (C) accepts the idea that staff behavior affects patients, (D) is aware of patients' needs, (E) perceives a need for staff congeniality, (F) views patients unfavorably, (G) thinks staff should be critical about performance.

This treatment-disposition behaviors evaluated were the (a) average hours of restraint and seclusion, (b) number of unauthorized absences and elopements of patients, (c) number of maximum hospital benefit (MHB) discharges, (d) number of patients released on trial visit, (e) number of irregular discharges, and (f) total dispositions including all releases and discharges. The analyses of these treatment-disposition behavior variables were based on the monthly figures of the units for the first 2 yr. under the unit system, corrected for the average monthly patient census.

Procedures

The attitude questionnaire (POT) was administered to as many of the staff members as possible three times with 6 mo. between administrations. Since not all of the staff members who took the questionnaire took it all three times, an analysis was used which permitted maximum use of the available data. The mean score was computed for each individual who took the test whether he took the test once, twice, or three times. The unit mean of this measure was selected as the best estimate of the overall staff attitude of the units. The sample of

418 included 82 staff members from Unit I, 141 from Unit II, 95 from Unit III, and 100 from Unit IV. The average participation across all three administrations of the test was about 60% for Unit I, 85% for Unit II, 70% for Unit III, and 75% for Unit IV. The proportion of the different professions represented in the samples from each unit was comparable, and these proportions in each unit were comparable to the proportions of the professions in the total staff.

In order to focus attention on the more meaningful variables, staff attitude and treatment-disposition behavior measures were screened on their ability to discriminate among the units. Only measures showing significant differences among the units, or ratios of these measures, were analyzed in the main comparisons relating staff attitudes to each other and to treatment-disposition behavior measures at the unit level.

RESULTS AND DISCUSSION

Selection of Staff Attitude and Treatment-Disposition Behavior Measures

Results of analyses of variance screening the seven attitude areas on the ability to discriminate between units are shown in Table 1. Statistically significant differences were found between the units in the following areas: interprets rules in an authoritarian manner ($p < .01$), thinks patients should be told their condition ($p < .001$), accepts the idea that staff behavior affects patients ($p < .01$), and views patients unfavorably ($p < .05$).

.01), and views patients unfavorably ($p < .05$).

Table 2 presents the results of analyses of variance screening the treatment and disposition behaviors on ability to differentiate units. Significant differences between units were found in the number of MHB discharges ($p < .001$), the average hours of restraint and seclusion ($p < .01$), and the number of trial visits ($p < .05$). The differences among the units in MHB discharges and trial visit releases did not seem to be the result of differences in total number of dispositions, as the units did not differ significantly on this variable ($F < 1$, $df = 3/69$).

Relationships among Personnel Attitudes and Treatment-Disposition Behavior Measures

Only those staff attitude and treatment-disposition behavior measures which indicated statistically significant differences among units or ratios of these measures were further evaluated. Since there were only four units, a very strong relationship be-

TABLE 2

TREATMENT-DISPOSITION BEHAVIOR MEASURES
AMONG THE UNITS

Variable	Source	df	MS	F
Hr. of restraint and seclusion	Units	3	19.77	4.41**
	Mo.	23	8.50	1.90*
	Residual	69	4.48	
No. of patients absent without official leave	Units	3	0.83	<1
	Mo.	23	0.20	<1
	Residual	69	1.06	
No. of trial visit releases	Units	3	3.93	2.75*
	Mo.	23	3.08	2.15*
	Residual	69	1.43	
No. of maximum hospital benefit discharges	Units	3	17.63	26.71***
	Mo.	23	4.56	6.91***
	Residual	69	0.66	
No. of irregular charges	Units	3	15.30	1.15
	Mo.	23	12.56	<1
	Residual	69	13.35	
Total no. of discharges and releases	Units	3	1.63	<1
	Mo.	23	9.23	1.55
	Residual	69	5.95	

Note.—The 24 monthly measures of all six variables were corrected for the average monthly patient census for each of the four units.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 1

STAFF ATTITUDES TOWARD PATIENT CARE
AMONG THE UNITS

Attitude Area	Source	df	MS	F
A. Interprets rules in an authoritarian manner	Units	3	128.67	4.29**
	Within	414	29.99	
B. Thinks patients should be told their condition	Units	3	89.33	7.08***
	Within	414	12.61	
C. Accepts the idea that staff behavior affects patients	Units	3	24.67	5.54**
	Within	414	4.45	
D. Is aware of patients' needs	Units	3	31.00	1.32
	Within	414	23.44	
E. Perceives a need for staff congeniality	Units	3	26.00	1.07
	Within	414	24.22	
F. Views patients unfavorably	Units	3	78.11	2.60*
	Within	414	30.10	
G. Thinks staff should be self-critical about performance	Units	3	0.17	<1
	Within	414	0.32	

Note.—The average scores for all individuals who took the Philosophy of Treatment Questionnaire at least once were used in the analyses.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

tween variables was required to attain statistical significance. Nonetheless, a number of significant relationships were found. All of the correlations between unit means on Areas A, B, and F were positive and statistically significant, indicating that units high on one variable were high on all three. Correlations of .99 ($p < .01$), .98 ($p < .02$), and .99 ($p < .01$) were found for the relationships between unit means of Areas A with B, A with F, and B with F, respectively. These results are shown in Table 3. As indicated above, the N was 4 for all these correlations. A strong belief that a patient should be informed of his psychiatric condition may be surprising in those units which endorse more authoritarian rule interpretation and less favorable attitudes toward patients.

More authoritarian interpretation of rules and less favorable views of patients are certainly not in vogue, and groups of people willing to endorse such views on a questionnaire could logically be considered relatively frank and open. When these same groups of people endorse attitudes of informing patients about their condition, they are endorsing openness and frankness in interactions with patients. Thus, a possible common thread among the three empirically related areas

would seem to be a generalized open and frank approach.²

The hypothesis that relationships among these area scores resulted from a generalized orientation of frankness and openness was evaluated further by correlating these with other variables. Openness and frankness are likely to be found in conjunction with trust of others. Specifically, if the endorsement of these three attitude areas (Areas A, B, and F) by unit staffs does indicate openness, frankness, and trust, then these attitudes in a unit atmosphere should be negatively correlated with hours of restraint and seclusion of patients. If trust is not indicated, positive

² The relationships of Areas A, B, and F in the unit comparisons might have been the result of a generalized response bias among the individuals. However, further analysis indicated that this was not the case. There were nine possible correlations (three on each of the three occasions the POT was administered to the staff) when Areas A, B, and F could be related, using a comparison of individuals rather than units. None of these nine correlations was significantly positive. On the contrary, only two were positive and the average of the nine was $-.07$. Thus, the strong positive relationships between Areas A, B, and F in the unit comparisons were not simply a reflection of the response tendencies of the individuals involved, but appeared to be a unit-level treatment-atmosphere phenomenon.

TABLE 3

CORRELATIONS OF UNIT MEANS OF STAFF ATTITUDE AND TREATMENT-DISPOSITION BEHAVIOR MEASURES

	1	2	3	4	5	6	7	8
1. Area A score ^a Authoritarian attitudes	—							
2. Area B score ^a Informing patients	.993**	—						
3. Area C score ^a Staff behavior importance	.936	.932	—					
4. Area F score ^a Unfavorable views of patients	.982*	.994**	.953*	—				
5. Hour of restraint and seclusion ^b	-.407	-.319	-.154	-.239	—			
6. Trial visits ^b	-.631	-.561	-.775	-.579	.296	—		
7. Maximum hospital benefits (MHB) discharges ^b	.692	.772	.600	.776	.092	-.106	—	
8. MHB/MHB + trial visits	.953*	.953*	.996**	.971*	.166	-.726	.655	—
9. % of staff participation	.760	.792	.919	.884	.251	-.593	.642	.911

^a The average scores of all individuals who took the Philosophy of Treatment Questionnaire at least once were used.

^b These measures were corrected for average patient census.

* $p < .05$.

** $p < .01$.

relationships would be anticipated. The correlations with hours of restraint and seclusion were $-.41$, $-.32$, and $-.24$ for Attitude Areas A, B, and F, respectively. Although not statistically significant, these negative correlations tend to support the trust interpretation.

Trust might also be reflected in the conditions under which patients were released from the hospital. A trial visit release does not sever the patient-hospital relationship and facilitates rehospitalization, usually indicating less trust in the patient's capabilities. An MHB discharge severs the patient's relationship with the hospital, ordinarily indicating greater trust in his capabilities.^a The ratio of MHB discharges to the sum of MHB discharges plus trial visits in a unit was used as the index of trust of patients' capabilities. When the mean scores of the units on Areas A, B, and F (the attitudes hypothesized to indicate a frank, open, and trusting orientation) were related to this measure of trust, there were significant positive relationships between all three attitude areas and the trust measure. The correlations of the areas with the mean of the ratio of number of MHB discharges divided by the sum of the MHB discharges plus trial visits were $.95$ ($df = 2$, $p < .05$) for Areas A and B, and $.97$ ($df = 2$, $p < .05$) for Area F. These statistically significant findings also indirectly support the openness, frankness, and trust interpretation of the relationships found between this group of endorsed attitudes of the staff toward patient care.

The frankness, openness, and trust dimension interpreted from the unit-level treatment-atmosphere data of the present study could be viewed as the antithesis of the pro-

tective benevolence factor for individuals found by Ellsworth (1965). What he called protective benevolence attitudes toward patients were the endorsed beliefs that patients should be made to feel at home and be approached in a generally nonpunitive manner. However, patients perceived people who endorsed these attitudes as aloof, distant, and noninteracting individuals. Ellsworth's interpretation of the discrepancy was that the high-protective-benevolence individual prefers to "go along with" patients and to accept them as they are without any attempt to change them. Thus, confrontations (disagreeing with patients' statements or demands) and the potential stresses and differences of opinion that could result from confrontations can be avoided. Ellsworth concluded that the high-benevolence individual apparently strives for superficially comfortable relationships with patients by accepting them as they are and avoiding confrontations.

In contrast, the frankness, openness, and trust orientation in the present study involves acceptance of a more authoritarian and less laissez-faire approach (higher Area A scores), less acceptance of patients as they are (less favorable ratings of patients, Area F scores), but greater trust in their abilities to handle stress (stronger belief that patients should be informed about their condition, Area B scores and a higher MHB discharge to trial visit release ratio). Thus, it is possible that the proposed openness, frankness, and trust orientation might also include an orientation toward helping patients change.

The different rates of participation of staff in the units indicated a differential level of cooperation among staffs in the units. The high positive (albeit not significant due to the small number of units) correlations between the units' percentage of staff participation and their endorsement of the Attitude Areas A, B, and F (Table 3) support the interpretation that high scores by units on these areas indicate trust of others. That the units high in scores on these areas tend to trust others more was indicated by the willingness of more people to expose their attitudes.

Although differential participation would

^a The possibility that MHB discharges which sever the patients' relationship with the hospital were actually a desire to get rid of the patients rather than a trust in the patients' capabilities was evaluated. When MHB discharges were related to irregular discharges on a unit level, the correlation was significantly negative ($r = -.92$, $p < .05$). As irregular discharges separate patients from the hospital and preclude readmission for 90 days, the negative relationship between MHB and irregular discharges was interpreted as support for the hypothesis that MHB discharges are indicative of trust rather than a desire to get rid of patients.

usually necessitate greater caution in interpretation of results, this did not seem to be the case in the present study. First, differential participation did not result in inequalities in professional representation in the samples from the units, indicating that this potential confounding was not present.

In addition, differential participation probably *reduced* the strength of the critical correlations between staff attitudes and treatment-disposition behavior by restricting the range of the attitude means for the units: (a) It would seem likely that nonresponders would tend to be low in frankness, openness, and trust of others; (b) complete returns from all units would be expected to spread the unit means of the frankness, openness, and trust of others attitudes, since the units with the fewest respondents tended to have lower means on these attitudes; and (c) the actual range of unit-staff attitude score means was probably reduced. Reduced range of attitude means would result in lower correlations between staff attitudes and the behavioral measure of trust of patients' capabilities, thereby decreasing, rather than increasing, the probability of falsely rejecting the null hypothesis.

In light of the findings above, the significant positive relationships between all three measures in the openness, frankness,

and trust cluster of endorsed attitudes and the objective measure of trust of patients would seem to indicate that openness, frankness, and trust could be important dimensions in the treatment climate and merit further exploration. Specifically, evaluation of the relationships between these dimensions and an orientation toward helping patients change might be fruitful.

REFERENCES

- BARRELL, R. P., DeWOLFE, A. S., & CUMMINGS, J. W. A measure of staff attitudes toward the care of physically ill patients. *Journal of Consulting Psychology*, 1965, 29, 218-222.
- CANTER, F. M. The relationship between authoritarian attitudes, attitudes toward mental patients, and effectiveness of clinical work with mental patients. *Journal of Clinical Psychology*, 1963, 19, 124-127.
- CARSTAIRS, G. M., GILBERT, D. C., HERON, A., LEVINSON, D. J., & PINE, F. Ideology, personality, and role definition: Studies of hospital personnel. In M. Greenblatt, D. J. Levinson, & R. H. Williams, (Eds.), *The patient and the mental hospital*. Glencoe, Ill.: Free Press, 1957.
- COHEN, J., & STRUENING, E. L. Opinions about mental illness: Hospital social atmosphere and patient time in hospital. *Journal of Consulting Psychology*, 1964, 28, 291-298.
- ELLSWORTH, R. B. A behavioral study of staff attitudes toward mental illness. *Journal of Abnormal Psychology*, 1965, 70, 194-200.

(Received August 2, 1967)

EFFECTS OF LABELS ON THE PERCEPTION AND DISCRIMINATION LEARNING OF RETARDATES¹

PHYLLIS A. KATZ AND SUE ROSENBERG

New York University

The role of labels in perception was assessed in 54 familial retardates and 48 normal IQ children at two matched MA levels. Three experimental manipulations were employed at each level in a factorial design. One group learned to associate a common label to random visual forms; a second group learned distinctive labels to the same forms; finally, a third group viewed the forms without labels. Following verbal training, all Ss received perceptual judgment and discrimination-learning tasks. No significant differences emerged between the retarded and normal Ss on paired-associate learning, discrimination learning, or perceptual accuracy without labels. In contrast to the normal children, however, the retardates' perceptual and learning processes were relatively unaffected by the learning of verbal levels.

A recent controversy in the area of mental retardation has been centered around the question of whether or not language cues have the same effect on normal children and familial retardates. One widely accepted theoretical position advanced by Luria (1963) suggests that the cognitive and motor behavior of retarded children is not effectively regulated by language. This relative lack of verbal mediation is assumed to result from underlying neurological defects. A second view, espoused by Zigler (1966), maintains that familial retardates are not linguistically deficient, but are instead intellectually comparable to normal children of the same MA level. This latter position holds that differences in intellectual performance previously obtained between retarded Ss and matched MA normals (e.g., Ellis, 1963; Kounin, 1941; Spitz, 1963; Zeaman & House, 1963) are attributable to differential motivational patterns rather than to innate organismic defects.

Research relevant to these two positions has focused primarily upon the effects of

verbal training on learning and problem solving in retardates (e.g., Barnett & Cantor, 1957; Milgram & Furth, 1963; O'Conner & Hermelin, 1959; Smith & Means, 1961), and the findings have not provided clear-cut support for either viewpoint. Despite the widely acknowledged importance of perception as the process underlying learning and problem solving, there have been no investigations, to date, of the direct effects of language upon the perceptual processes of retarded individuals. The purpose of the present study was to compare the effects of verbal training upon the perception of retardates and matched MA normals. Based upon the reasoning outlined above, Luria's (1963) position would generate the prediction that the perceptual responses of retardates will be less influenced by language training, whereas Zigler's (1966) viewpoint would predict no difference in the degree of modification exhibited by the two groups. These two possibilities were assessed in the present investigation.

METHOD

Subjects

Fifty-four familial retardates were selected from a New York State public institution for the mentally retarded. Half of the Ss had a mean MA of 6 yr., 10 mo. on the Stanford-Binet Intelligence Scale at the time of the experiment, and the other half had an MA of 8 yr., 9 mo. These groups will be referred to hereafter as the low- and the high-MA groups, respectively. The mean IQ of these Ss were 47.7 and 60.2, respectively, with a total range extending from 40 to 73. The mean CA of

¹ This study was partially supported by a New York University Arts and Science Research Grant and by National Institute of Mental Health Grants HD-2672-01, Phyllis A. Katz, principal investigator and HD-MH-03008, Edward Zigler, principal investigator. The authors would like to thank Edward Zigler for assisting in obtaining Ss and Jacob Schneider and Thomas McCulloch for their assistance in providing them. Additionally, particular gratitude is expressed to Aron B. Katz for his many insightful suggestions.

both the groups was 19 yr., 6 mo. This sample comprised all of the higher MA retardates available at the institution. No *S* whose records indicated any evidence of organic etiology was included in the sample.

The performance of the retarded *Ss* was compared to scores obtained by a group of 48 *Ss* of normal intelligence in an earlier study (Katz, 1963). The normal *Ss* were randomly drawn from the first-, second-, and fourth-grade classes of a public elementary school in New Haven, Connecticut. Half of these *Ss* were 7 yr. old, and the other half averaged 9 yr., 6 mo. Thus, the normal and retarded *Ss* were of roughly comparable MA levels. It was not possible to match the higher level retarded *Ss* exactly to the older normal sample (who were run earlier) since there were not enough retardates at the institution with MA levels of 9 yr. or more.

Procedure

All *Ss* were run individually in the school or institution. The tasks were introduced to *Ss* as games. Upon entering the experimental room, they were shown "prizes" and were told they could win the object of their choice if they played the games well. The prizes were appropriate to the CA of *Ss*. For the normal elementary school children, the reinforcements included such things as jacks, modeling clay, toy guns, marbles, books, and small games. For the retardates, the prizes included such objects as comic books, combs, stationery, and ball-point pens.

Within the retarded and normal groups, at both MA levels, an equal number of *Ss* was randomly assigned to three experimental condition, depending upon the type of verbal training received. A common label group was taught to associate two nonsense syllables with four highly similar random forms. Thus, for this group, there were two forms per label. A distinctive label group learned to associate four distinctive labels to the same four forms. Finally, a control group viewed the figures without labels. The control *Ss* were asked to count the figures aloud to minimize the possibility of spontaneous labeling. For all training conditions, the forms were individually presented on slides for 150 trials. A more detailed description of the stimuli and the apparatus can be found elsewhere (Katz, 1963). The verbal labels were chosen from nonsense-syllable lists of 95% association value and were randomly assigned to the visual forms for each *S*. A correction procedure was used for the label groups. Verbal support was given to all groups after every tenth trial, that is, "You're doing fine"; "You really know how to play this game"; etc. In order to maintain interest in the task, the nonsense forms were introduced to all *Ss* as "pictures of funny-looking animals from Mars," and *Ss* were asked to pretend they were space explorers.

Following the verbal training, a perceptual judgment task was introduced in which pairs of the previously employed forms were tachistoscopically presented on slides at a .2-sec. exposure. The *S* was

instructed to make a judgment of "same" if the members of the pair looked like twins (i.e., were identical) and "different" if they did not. Judgments were indicated by *S*'s depression of one of two appropriately marked telegraph keys, which stopped a clock when pressed. The apparatus allowed for the recording of both choice scores and judgment latencies. The slides which were shown contained 8 pairs of identical forms and 20 pairs of two different forms. Ten slides in this latter group contained pairs of different figures previously associated with the same label for the common label group. The actual forms used were matched across groups.

Following the perceptual judgment task, a discrimination-learning task was introduced to all *Ss*. Three of the four previously employed forms were used as discriminanda. For this task, the forms were cut out of Masonite and mounted on blocks of wood. The *S* was instructed as follows: "One of the animals has a paper goldfish like this under it. You pick up the one you think has the fish under it. If you're right, you can keep the fish and at the end, you can trade them in for prizes."

The reinforced stimulus was always one of the two forms previously associated with the same label for *Ss* in the common label group. The position of the figures was changed behind a vertical screen before each trial. Testing was discontinued after five consecutive correct choices or 50 trials.

RESULTS

Verbal Learning

A four-way analysis of variance (Type of Label \times Type of *Ss* \times MA Level \times Blocks of Trials) was conducted on the number of correct responses made by the various label groups. The only significant ($p < .01$) effect to emerge from the analysis was the blocks of trials, indicating that all *Ss* improved over trials. None of the other main or interaction effects was significant. The normal and retarded *Ss* at each MA level learned the labels equally well.

Perceptual Judgments

The measure of major interest to the positions initially advanced is the number of "same" judgments made to pairs of physically different forms which were associated with the same label in the common label group. It would be expected that if labels did modify perception, then a common name should elicit more "same" judgments made to the two figures, whereas distinctive names should elicit more "different" judgments relative to a no-label control group.

The mean number of "same" judgments made by the various groups is presented in Table 1. A three-way (Type of $S \times$ Treatment \times MA) analysis of variance was conducted on these scores. This analysis revealed the type of S effect to be significant ($F = 7.93$, $df = 1/90$, $p < .01$), indicating that the retarded Ss were somewhat more inaccurate in their perception than the normals. A comparison of the normal and retarded no-label control groups was not significant, however, suggesting that the type of S effect is primarily attributable to the labeling training. The Treatment $\times S$ interaction was significant at the .05 level ($F = 4.71$, $df = 2/90$), reflecting the finding that the verbal training conditions had differential effects on retarded and normal Ss. In addition, the triple interaction of Treatment $\times S \times$ MA level was significant ($F = 3.77$, $df = 2/90$, $p < .05$).

In view of these interactions involving type of S , separate analyses were conducted for the normal and retarded Ss. For the normals, the main effect of treatment was statistically significant ($F = 4.70$, $df = 2/42$, $p < .05$), indicating that labels influenced the judgment of stimulus similarity in the predicted direction. Analysis of the retardates' error scores, however, revealed no significant overall treatment effect. For the retarded group, a significant Treatment \times MA interaction emerged. The means involved in this interaction, contained in Table 1, reflect the finding that labels did not influence the judgments of the low-MA retarded Ss at all, whereas they did have some effect on the higher MA Ss. For

TABLE 1
MEAN NUMBER OF "SAME" JUDGMENTS MADE TO
PHYSICALLY DIFFERENT PAIRS OF FORMS ASSO-
CIATED WITH THE SAME LABEL FOR THE
COMMON LABEL GROUPS

Group	Type of subject			
	Normal		Retardate	
	Low MA	High MA	Low MA	High MA
Common label	4.00	4.62	5.43	4.44
No label (Control)	2.35	3.88	5.71	2.11
Distinctive label	2.25	1.62	5.33	5.55

TABLE 2
MEAN RESPONSE LATENCIES ON PERCEPTUAL
JUDGMENT TASK

Group	Type of subject	
	Normal	Retardate
Label	1.63	1.41
Control	1.18	1.57

this latter group, having an equivalent label for two stimuli made the stimuli appear less distinctive, as was true for the normals. Having distinctive labels, however, which theoretically should decrease similarity, had an opposite effect for the high-MA retardates.

The data regarding response latencies to these stimulus pairs paralleled the choice scores, for the most part. When the scores of normal and retarded Ss were analyzed together, a significant Treatment \times Type of S interaction emerged. The means involved in this interaction are presented in Table 2. It can be seen that with the normal children the label groups took longer to make a judgment than did the control group ($F = 4.88$, $df = 2/36$, $p < .05$), suggesting that verbal mediation was operative. For the retarded Ss, however, no such difference between the label and control groups in response latencies was evident.

Discrimination Learning

The mean number of correct responses is contained in Table 3. An analysis of variance conducted on these measures revealed a significant Treatment \times Type of S interaction ($F = 3.40$, $df = 2/90$, $p < .05$), indicating, once again, that the labels had differential effects on retarded and normal Ss. Analyses conducted separately on these two populations revealed a significant treatment effect

TABLE 3
MEAN NUMBER OF CORRECT RESPONSES ON
DISCRIMINATION-LEARNING TASK

Group	Normal		Retardate	
	Low MA	High MA	Low MA	High MA
Common label	15.5	17.1	23.0	21.0
No label (control)	19.4	23.1	20.1	30.1
Distinctive label	27.9	26.5	15.7	23.1

for the normal Ss. It can be seen in Table 3 that this effect is in the expected direction; that is, the common label group made more errors than a control group, whereas the distinctive label group made fewer errors. An analysis of variance conducted on the retardates' scores yielded no significant main or interaction effects. Thus, it may be concluded that labels did not influence the discrimination learning of the retardates.

DISCUSSION

The major finding of the present investigation is that labeling training has differential effects on the perceptual responses of familial retardates and normal children of matched MA levels. For normal 7- and 9-yr.-old school children, the association of a common label with two objects elicited increased perceptual similarity and more difficulty in learning to discriminate between them relative to a no-label control group. Conversely, having distinctive labels for these stimuli increased distinctiveness and facilitated the learning of a discriminatory response. Furthermore, groups having labels took longer to make perceptual judgments than groups without labels. This latter finding offers additional corroboration for verbal mediation effects upon perception. With the retarded population, however, labeling training did not significantly alter, for the most part, either the discriminability of the stimuli or the time required to make a perceptual response. This clearly suggests that verbal mediation training was not particularly effective in modifying the perception of retarded Ss.

Although it is clear from the findings that labels have less salience in the discriminatory processes of retarded Ss than in normals, it should be noted that several possible interpretations of this difference exist.

One possibility is that Luria's (1963) general thesis is correct, and that some as yet unspecified organic linguistic defect exists in all retardates. Logically, however, such an assumption could not be unequivocally supported in the absence of direct neurological evidence.

A second possible explanation is that verbal cues are less effective for the retardate be-

cause he has received less reinforcement for verbal mediation in the past. A caretaker's verbal interactions with a retarded child might be minimal, particularly under certain types of institutional conditions, and thus words may not acquire much secondary reinforcing value for the retardate. This social reinforcement interpretation suggests that verbal mediation differences would be less pronounced between a noninstitutionalized retarded sample and a normal group of children. Furthermore, it offers more potential for behavior change than the cognitive structural defect position postulated by Luria.

A third possibility which should be considered is that the retarded Ss were not verbally deficient at all. It is typically assumed by most investigators that a retardate behaving differently than a matched MA normal is necessarily functioning at a lower developmental level. Such an assumption completely ignores the greater opportunities for learning that the chronologically older retardate has had. Although the developmental trends regarding verbal mediation in normal individuals are not completely clear, extrapolation from a considerable body of research in this area suggests that the salience of verbal cues may be more pronounced in the perceptual processes of children (e.g., Cantor, 1955; Katz, 1963; Norcross, 1958) than of adults (e.g., Arnoult, 1953; H. C. Ellis & D. G. Muller, 1964; Robinson, 1955). Thus, it is conceivable that the retardates in this study were behaving more like their CA counterparts. Further testing is indicated to evaluate these various possibilities.

REFERENCES

- ARNOULT, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *Journal of Experimental Psychology*, 1953, **45**, 401-409.
- BARNETT, C. D., & CANTOR, G. N. Discrimination set in defectives. *American Journal of Mental Deficiency*, 1957, **62**, 334-337.
- CANTOR, G. N. The effects of three types of pre-training on discrimination learning in preschool children. *Journal of Experimental Psychology*, 1955, **49**, 339-342.
- ELLIS, H. C., & MULLER, D. G. Transfer of perceptual learning following stimulus predifferentiation. *Journal of Experimental Psychology*, 1964, **68**, 388-395.

- ELLIS, N. R. The stimulus trace and behavioral inadequacy. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.
- KATZ, P. A. Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology*, 1963, 66, 423-428.
- KOUNIN, J. Experimental studies of rigidity: I. The measurement of rigidity in normal and feeble-minded persons. *Character and Personality*, 1941, 9, 251-272.
- LURIA, A. R. Psychological studies of mental deficiency in the Soviet Union. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.
- MILGRAM, N. A., & FURTH, H. G. The influence of language on concept attainment in educatable retarded children. *American Journal of Mental Deficiency*, 1963, 67, 733-739.
- NORCROSS, K. J. Effects on discrimination performance of similarity of previously acquired stimulus names. *Journal of Experimental Psychology*, 1958, 55, 305-309.
- O'CONNOR, N., & HERMELIN, B. Discrimination and reversal learning in imbeciles. *Journal of Abnormal and Social Psychology*, 1959, 59, 409-413.
- ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *Journal of Experimental Psychology*, 1955, 49, 112-114.
- SMITH, M. P., & MEANS, J. R. Effects of type of stimulus pretraining on discrimination learning in mentally retarded. *American Journal of Mental Deficiency*, 1961, 66, 259-265.
- SPITZ, H. H. Field theory in mental deficiency. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.
- ZEAMAN, D., & HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.
- ZIGLER, E. Mental retardation: Current issues and approaches. In L. W. Hoffman & M. L. Hoffman (Eds.), *Review of child development research*. New York: Russell Sage Foundation, 1966.

(Received July 29, 1967)

Manuscripts Accepted for Publication in the
Journal of Abnormal Psychology

- Prediction of Reaction to Placebo: Michael M. Nash and Fred M. Zemring*: Counseling and Psychotherapy Research Center, 5846 University Avenue, University of Chicago, Chicago, Illinois 60637.
- Psychology, Metapsychology, and Psychoanalysis: Paul L. Wachtel*: Research Center for Mental Health, New York University, 4 Washington Place, New York 10003.
- Effect of Peer Model Responsiveness and Length of Induction Procedure on Hypnotic Responsiveness: Burton I. Klinger*: Department of Psychology, University of New Hampshire, Durham, New Hampshire 03824.
- Verbal Conditioning of Psychopaths as a Function of Experimenter-Subject Sex Differences: Denton J. Stewart and Jerome H. Resnick*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.
- Analysis of the Treatment Effects of Stimulation Instructions in the Application of the Real-Simulating Model of Hypnosis: Peter W. Sheehan*: University of New England, Armidale, New South Wales, 2351 Australia.
- Comparison of the Restricted Association of Chronic Schizophrenic and Normal Control Subjects: Elliot J. Stern and Klaus F. Reigel*: Language Development Program, 330 Nickels Arcade, Ann Arbor, Michigan 48108.

* Asterisk indicates author for whom the address is supplied.

A BRIEF MENTAL HEALTH INDEX¹

E. K. ERIC GUNDERSON AND RANSOM J. ARTHUR

*Navy Medical Neuropsychiatric Research Unit,
San Diego, California*

The Cornell Medical Index (CMI) was administered to 630 Navy psychiatric patients and 454 healthy controls. Patient and control samples were split into two groups for cross-validation purposes, and two methods, regression analysis and a new item selection technique called SEQUIN, were applied to the problem of selecting the most discriminating set of CMI items. The percentages correctly classified "sick" or "well" when results from Sample 1 were used to predict Sample 2 and vice versa were 82% and 85% by the regression method and 86% and 86% by the SEQUIN method. Seven items, perhaps representing general attributes defining mental illness in the Navy culture, contributed significantly to the predictive scales regardless of particular item selection method or sample.

The Cornell Medical Index (CMI) has been widely used as an aid to clinical diagnosis (Brodman, Erdmann, & Wolff, 1956; Erdmann, 1959) and as a measure of the prevalence of somatic and emotional symptoms in various populations (Abramson, Terespolsky, Brook, & Kark, 1965). The Index has repeatedly demonstrated validity as an indicator of general health or of emotional health when compared with clinical ratings, but has proved less useful as an indicator of the presence of specific diseases or in comparing health status in different populations and cultures (Abramson, 1966). In a previous study of the Navy population, the total CMI score (the number of "yes" responses over the entire 195-item health questionnaire) discriminated efficiently between psychiatric inpatients and healthy controls using a cutting score of 15 (Arthur, Gunderson, & Richardson, 1966). This result suggested that the CMI might have potential value as a psychiatric screening device and as an epidemiological tool in the Navy setting.

The next logical step in developing a useful screening index was to identify the specific

items in the CMI which best differentiate patients from healthy controls and to determine the most efficient combination of these items. It seemed plausible that a short scale of highly selected items might be devised which could discriminate mentally ill from healthy individuals in the Navy population as well as the entire CMI inventory.

METHOD

Subjects

The Ss for the study were 630 psychiatric inpatients from two naval hospitals and a normal control group of 454 enlisted men who were currently on active duty and were not medical or psychiatric patients at the time of the study. The control group consisted of students from an electronics technicians school, Marine guards, hospital corpsmen, and Navy enlisted men from several occupational categories on duty at a West Coast naval station, and participants in underwater demolition training. The mean ages of the patient group and the control group were 22 yr. and 21 yr., respectively. Distributions in pay grade (rank) were approximately the same for both patients and controls, although seamen recruits were represented in the patient group but not in the control group.

Procedure

CMI questionnaires were administered individually or in groups by medical personnel at the various facilities included in the study. The total sample of 1,084 patients and controls as divided into two subgroups for purposes of cross-validation by sorting on odd or even last digits of service numbers. By this method, 560 cases were selected for Sample I, of whom 335 were patients and 225 were controls. The composition of Sample II was 295 patients and 229 controls, for a total of 524 Ss.

¹ Report No. 67-18, supported by the Bureau of Medicine and Surgery, Navy Department, under Research Work Unit MF 022.01.03-9002. The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Navy.

Statistical assistance was provided by Frank Thompson.

The frequency of "yes" responses given by patients and controls for all 195 CMI items was examined, and, on the basis of the differences in percentages answering "yes," 60 items were chosen for intensive analyses. The regression technique was used to select the most discriminating of the 60 items and to determine regression weights for sets of items derived separately from Samples I and II. The regression weights from Sample I then were used to predict the patient versus control criterion in Sample II (0 = control, 1 = patient), and, similarly, regression weights from Sample II were used to predict the illness criterion in Sample I. An alternative method of item selection and weighting, called SEQUIN (Moonan & Pooch, 1966), also was used for cross-validation. The most important difference between the two methods is the assignment of unit weights to items selected by the SEQUIN method as opposed to the assignment of variable weights in the regression method.

Correlations were computed by the product-moment method for the cross-validities as well as the regression analysis and by the biserial method for the SEQUIN analysis; the two methods of correlation yielded identical results in the present analyses.

RESULTS

Table 1 presents the items, correlations with the criterion, and the regression coeffi-

cients derived from Samples I and II. The items most highly correlated with the criterion in both samples were, "Does worrying continually get you down?" and "Are you considered a nervous person?" The item, "Were you ever a patient in a mental hospital (for your nerves)?" had relatively large regression weights in both samples. Of all CMI questions, these items most directly refer to severe emotional disturbance. Substantial weights were received by the items, "Do you smoke more than 20 cigarettes a day?" "Do you suffer badly from frequent severe headaches?" and "Do you need glasses to read?" The relationship of these items to mental illness is less obvious than the relationships of the aforementioned items; however, headaches, reading difficulties, and excessive smoking are commonly recognized concomitants of emotional distress. Various specific emotional symptoms, for example, "shy or sensitive," "frightening thoughts," and "unhappy and depressed" contribute to discrimination

TABLE 1

CORRELATIONS AND REGRESSION COEFFICIENTS FOR CORNELL MEDICAL INDEX (CMI)
ITEMS SELECTED BY THE REGRESSION METHOD

Item no.	CMI section	Item	Sample I		Sample II	
			r^a	Regression coefficient	r	Regression coefficient
1	A	Do you need glasses to read?	.23	.13	.27	.16
79	G	Do you suffer badly from frequent severe headaches?	.41	.13	.38	.14
113	I	Do you suffer from severe nervous exhaustion?			.39	-.11
133	K	Are you definitely underweight?	.30	.12	.32	.11
142	L	Do you smoke more than 20 cigarettes a day?	.42	.23	.39	.13
143	L	Do you drink more than six cups of coffee or tea a day?			.30	.12
146	M	Do you get nervous and shaky when approached by a superior?	.40	.13		
158	N	Do you usually feel unhappy and depressed?	.45	.14		
162	N	Do you often wish you were dead and away from it all?			.40	.16
163	O	Does worrying continually get you down?	.48	.09	.49	.18
166	O	Are you considered a nervous person?	.47	.16	.46	.14
168	O	Did you ever have a nervous breakdown?			.35	.13
169	O	Did anyone in your family ever have a nervous breakdown?	.20	-.08		
170	O	Were you ever a patient in a mental hospital (for your nerves)?	.28	.21	.32	.19
172	P	Are you extremely shy or sensitive?	.36	.13		
180	Q	Are you easily upset or irritated?	.39	-.11		
184	Q	Do people often annoy and irritate you?			.45	.11
189	R	Do sudden noises make you jump or shake badly?	.43	.10		
193	R	Do frightening thoughts keep coming back in your mind?	.42	.13		
194	R	Do you often become suddenly scared for no good reason?	.28	-.13		
		Multiple r	.69		.69	

Note.—Items which contributed significantly ($p < .05$) to an increase in the multiple correlation were included.
^a Correlation between item responses (yes-no) and patient versus control criterion.

TABLE 2
CORRELATIONS OF ITEMS SELECTED BY THE SEQUIN METHOD

Item No.	Cornell Medical Index section	Item	Sample I	Sample II
1	A	Do you need glasses to read?	.22	.27
79	G	Do you suffer badly from frequent severe headaches?	.40	.38
85	G	Have you fainted more than twice in your life?	.28	
133	K	Are you definitely underweight?	.30	.32
139	L	Do you usually have great difficulty in falling asleep or staying asleep?		.45
142	L	Do you smoke more than 20 cigarettes a day?	.42	.39
143	L	Do you drink more than six cups of coffee or tea a day?	.31	.30
146	M	Do you get nervous and shaky when approached by a superior?	.40	
147	M	Does your work fall to pieces when the boss or a superior is watching you?	.40	
158	N	Do you usually feel unhappy and depressed?	.44	
162	N	Do you often wish you were dead and away from it all?		.39
163	O	Does worrying continually get you down?	.48	.49
166	O	Are you considered a nervous person?	.46	.45
167	O	Does nervousness run in your family?		.29
168	O	Did you ever have a nervous breakdown?		.35
170	O	Were you ever a patient in a mental hospital (for your nerves)?	.28	.32
175	P	Does criticism always upset you?		.37
184	Q	Do people often annoy and irritate you?		.44
189	R	Do sudden noises make you jump or shake badly?	.43	
193	R	Do frightening thoughts keep coming back in your mind?	.42	
		Cumulative r	.67	.68

Note.—The correlations are between item responses (yes-no) and patient versus control criterion.

in one or the other sample, but are less consistent than the previously mentioned items.

Three suppressor items appeared in the Sample I analysis and one in the Sample II analysis. These items correlated positively with the criterion, but in combination with the entire set of significant items they received negative regression weights for the prediction of the patient-control criterion. The fact that none of the negative (suppressor) weights were replicated in the other sample would recommend caution in interpreting these results.

Items selected by the SEQUIN method and their correlations with the criterion for Samples I and II are shown in Table 2. All but five of the items selected by the SEQUIN method also had been selected by the regression method in at least one sample. Thus, the item content of the scales derived by the two methods was quite similar.

Multiple or cumulative correlations and cross-validation results are shown in Tables 3 and 4. Fifteen items significantly contributed to discrimination of patients from controls in Sample I by the regression method.

A multiple correlation of .69 was achieved with this set of items against the illness criterion. When the regression weights in Sample I were used to predict "well" versus "sick" status for Sample II, a cross-validity coefficient of .69 was obtained. Regression weights for 12 significant items were derived from Sample II in order to predict patient-control status for Ss in Sample I. The multiple-

TABLE 3
MULTIPLE CORRELATIONS AND CROSS-VALIDITY COEFFICIENTS FOR TWO ITEM-SELECTION METHODS

Method	No. of items	Multiple correlation	Cross-validation ^a
Regression Sample I Sample II	15	.69	.69
	12	.69	.64
		Cumulative correlation	
SEQUIN Sample I Sample II	14	.67	.72
	14	.68	.71

^a Items and weights from Sample I were used to predict patient or control status in Sample II and vice versa.

TABLE 4
CROSS-VALIDITY DISTRIBUTIONS FOR
TWO ITEM-SELECTION METHODS

Method	Sample I		Sample II	
	Actual		Actual	
	Well	Sick	Well	Sick
Regression Predicted ^a				
Sick	47	259	42	244
Well	170	45	180	34
Correct predictions	82%		85%	
SEQUIN Predicted ^a				
Sick	39	295	38	261
Well	186	40	191	34
Correct predictions	86%		86%	

^a Predictions for Sample I were based upon items and weights derived from Sample II and vice versa. Cutting points for predicted criterion scores were based upon actual proportions "sick" and "well" in each sample.

correlation and cross-validity coefficients were .69 and .64, respectively, for this set of items.

Using the SEQUIN technique for item selection, 14 items were found to cumulatively increase the correlations with the criterion for both Samples I and II. Cumulative correlations of these sets of items with the illness criterion were .67 and .68, respectively, and cross-validity correlation coefficients were .72 and .71, respectively, for Samples I and II.

The distributions obtained when the regression or SEQUIN results from Sample I were used to predict Sample II and vice versa are given in Table 4. The percentages classified correctly, that is, "sick" or "well," were 82% and 85% by the regression method and 86% and 86% by the SEQUIN method.

DISCUSSION

The development of a short, self-administered mental health index which provides a reasonable approximation of the results of a brief psychiatric examination would have obvious importance for both research and practice. Opportunities for comparative studies of populations using clinical examination methods are extremely limited, and investigations of prevalence of symptoms and ill-

ness almost always must rely upon Ss' own reports of health status. Important relationships of personality, cultural, and environmental factors to emotional illness and invalidism can be readily adduced by means of survey methods if valid questionnaire techniques can be devised.

In military psychiatric practice, decisions concerning fitness for military duty or the need to hospitalize often must be made rapidly and without the benefits of extended clinical study. A standardized screening instrument obviously could have value in settings with rapid patient turnover, for example, an outpatient clinic, where allocation of the psychiatrist's or physician's time is an important consideration. Also, the specific questions identified as most significant in differentiating "sick" from "well" individuals may suggest particularly useful areas to be explored in brief clinical interviews because these items have a clearly demonstrated relevance for estimating the probability of mental illness.

The findings of the present study are viewed as encouraging with respect to the possibility of developing a brief but efficient mental health index for the Navy population. It is clear that selection of items by linear multiple regression or similar techniques can provide great economies in scoring the present CMI questionnaire and at the same time slightly increase validity as well. Scoring only from 12 to 15 items by means of regression weights or unit weights resulted in concurrent validities consistently higher than those attained with scores based on unit weights for all 195 items in a previous study.² Split-half cross-validation results indicated high stability and comparability for the various brief predictor scales obtained.

The apparent superiority of the SEQUIN method of item selection for scale cross-validations suggests that simple counting of diagnostic answers ("yes") would be preferable

² When scores based upon all 195 items were dichotomized at 15 and were correlated with a sick-well criterion, a concurrent validity coefficient (ϕ) of .65 was attained. Details concerning the discriminating power of scores based upon the entire 195-item CMI were given in the study by Arthur et al. (1966).

to more complex differential weighting procedures in scoring a mental health index. This simplicity of scoring would offer practical advantages for many clinical settings.

A large proportion of the items included in the various experimental scales was drawn from Sections M-R of the CMI questionnaire. These items were designed to elicit psychological symptoms, and it is not surprising that they are the most discriminating with respect to mental illness. Three items from Section L—items usually not scored as emotional symptoms—also were highly significant correlates of diagnosed mental disorder. These items—"Do you usually have great difficulty in falling asleep or staying asleep?" "Do you smoke more than 20 cigarettes a day?" and "Do you drink more than six cups of coffee or tea a day?"—should be included in scales designed to estimate mental health status in the naval population.

Seven "universal" items emerged from the various analyses; that is, items which contributed significantly to the predictive scales regardless of particular item-selection method or sample. These questions perhaps represent a set of general attributes which are most important in defining mental illness in the Navy culture. The items, in order of importance, are, "Does worrying continually get you down?" "Are you considered a nervous person?" "Do you smoke more than 20 cigarettes a day?" "Do you suffer badly from frequent severe headaches?" "Are you definitely underweight?" "Were you ever a patient in a mental hospital (for your nerves)?" "Do you need glasses to read?" These items appear to offer a useful nucleus of questions for a brief psychiatric index. Other questions could be added as their relevance was consistently demonstrated.

Using different methods of item selection, Abramson et al. (1965) devised a scale of 10 "key" CMI items (Hebrew version) which

correlated highly ($r = .63$) with physicians' ratings of emotional health in a randomly selected sample of Jerusalem males. Seven of these 10 key items appeared in the experimental scales of the present study, suggesting a surprising generality in the present findings.

Studies of the relationships of brief mental health index scores to psychiatrists' decisions concerning fitness for duty and need for hospitalization are in progress, and the effects of social and cultural differences upon health-questionnaire responses also are being evaluated.

Because the predictive value of the brief, CMI scales derived in this study has not yet been tested, the accuracy and usefulness of these scoring techniques for diagnostic purposes with new or suspected patients remain to be demonstrated, and they obviously should not be used for intake screening until such verification can be achieved.

REFERENCES

- ABRAMSON, J. H. The Cornell Medical Index as an epidemiological tool. *American Journal of Public Health*, 1966, 56, 287-298.
- ABRAMSON, J. H., TERESPOLSKY, L., BROOK, J. G., & KARK, S. L. Cornell Medical Index as a health measure in epidemiological studies—a test of the validity of a health questionnaire. *British Journal of Preventive Social Medicine*, 1965, 19, 103-110.
- ARTHUR, R. J., GUNDERSON, E. K. E., & RICHARDSON, J. W. The Cornell Medical Index as a mental health survey instrument in the naval population. *Military Medicine*, 1966, 131, 605-610.
- BRODMAN, K., ERDMANN, A. J., JR., & WOLFF, H. G. *Medical index—health questionnaire manual*. New York: Cornell University Medical College, 1956.
- ERDMANN, A. J., JR. Experiences in use of self-administered health questionnaire. *American Medical Association Archives of Industrial Health*, 1959, 19, 339-344.
- MOONAN, W. J., & POOCH, U. W. *SEQUIN: A computerized item selection procedure*. (Res. Memo. SRM 67-8) San Diego, Calif.: United States Naval Personnel Research Activity, October 1966.

(Received July 26, 1967)

PERFORMANCE OF RETARDED READERS ON THE WISC AND EDUCATIONAL TESTS

J. G. LYLE AND JUDITH GOYEN¹

University of Sydney, New South Wales, Australia

The WISC results of 54 retarded readers (experimentals) of average intelligence from the six primary school grades were compared with those of a matched group of 54 adequate achievers (controls). The experimentals had greater Verbal (V) ~ Performance (P) differences than the controls in both directions. This was due mainly to their significantly poorer performances on Information, Arithmetic, and Coding subtests, and to their significantly better performances on Comprehension, Picture Arrangement, and Block Design subtests after adjustments had been made by covariance analyses for Full Scale IQ differences. The experimentals also performed significantly less well than controls on untimed tests of Spelling and Arithmetical Calculation. No tendency was found for the differences between groups on the WISC or educational tests to increase with age, as was the case with their reading achievement. Hence there was no support for Neville's hypothesis that certain of the differences are *effects* of reading retardation rather than correlated *symptoms*.

The concept of reading retardation as a symptom of minimal cerebral dysfunction (Money, 1962) has led psychologists to seek correlates of this disability in cognitive and perceptual tests. As with known cases of brain damage, retarded readers are found with gross Verbal ~ Performance discrepancies on the WISC or sometimes with gross discrepancies on just one or two of the WISC subtests. For example, the authors have found one child² of 8 yr. who averaged a scale score of 8 on all subtests excepting Vocabulary, on which he scored 0; and another (see Footnote 2) aged 17 yr. who, despite an overall IQ of 94, could not, even with prompting, attain a raw score of above 3 on Picture Completion.

In the clinic situation it is very tempting to infer correlations between such discrepancies and reading disability; but the clinical psychologist labors under two serious handicaps: he lacks a control group, and he is dealing with a selected clinic sample.

This report is part of a wider study designed to determine: (a) whether Verbal ~ Performance discrepancies on the WISC are characteristic of retarded readers; (b)

whether retarded readers manifest any characteristic scatter on the WISC subtests; and (c) whether reading retardation is a specific educational problem or whether it is merely part of a generalized school backwardness.

A number of studies of WISC patterns of retarded readers has been published, most of them suffering from the defects of small sample size ($N < 30$) and/or from lack of a control group. Studies falling within one or other of these categories are those of Altus (1956), Burks and Bruce (1955), Coleman and Rasof (1963), Dockerell (1960), Hirst (1960), Graham (1952), and Robeck (1962).

Experimentally more satisfactory studies are those of Neville (1961), McLeod (1965), and Belmont and Birch (1966), the two last mentioned being the only ones with large samples. Briefly, the findings in all three of these more satisfactory studies were that retarded readers perform relatively poorly on the subtests of Information and Arithmetic. Vocabulary was also found to discriminate significantly by McLeod (1965) and Belmont and Birch (1966), Digit Span by McLeod (1965) and Neville (1961), and Coding by McLeod (1965). Reverse discrimination (i.e., retarded readers better than controls) was found on the following subtests: Picture Completion (McLeod, 1965), Object Assembly (Belmont & Birch, 1966), and Picture

¹ The cooperation of the New South Wales Department of Education, in particular, of O. R. Jones, former Director of Primary Education, and C. H. von Sturmer, Principal Research Officer, is gratefully acknowledged.

² Not included in these samples.

Arrangement and Block Design (Neville, 1961).

Most published studies have drawn samples from specialized reading clinics, in which, because of their selective nature, factors of a sensory, neurological, emotional, or socio-economic nature are likely to have been unduly represented. This sampling bias has usually been compounded by a failure to examine Ss clinically in order to screen out those with easily explicable causes, such as culturally poor home backgrounds, histories of broken schooling, or even simple dullness. For example, McLeod (1965) may be criticized for the selection of both of his groups since he compared children from an Australian reading clinic (apparently without any screening of the individuals selected from the files) with a group of successful readers in England, who, because of differences in the educational systems, would have had a year's longer tuition than the experimental group. Belmont and Birch (1966), although selecting their samples from state schools, included in their retarded group 60 children who were under IQ 90, as against only one such normal reader. Not surprisingly, the discrepancy between the two groups in Full Scale IQ was accompanied by significant differences in all but one of the subtests. Equating the groups for IQ resulted in two attenuated samples of only 22 each.

But the major defect with all published studies is that they fail to define adequately the nature of reading retardation. The definition proposed here is that reading retardation is a *syndrome of learning disability in children who are bright enough to learn, and who have had the usual learning opportunities*. There is obviously no point in inquiring why dull children, or children with sensory handicaps, or children who have missed schooling, have progressed less rapidly in reading than their fellows.

The term "reading retardation" may in fact have four meanings: (a) slow reading speed, (b) limited vocabulary, (c) failure to master basic processes such as letter recognition and sound blending, and (d) lack of comprehension despite adequate fluency in the mechanics of reading. If *d*, termed "true dyslexia," is ignored as a rare con-

dition,^a it is *c* which should form the basis of the selection criteria; *a* and *b* can arise simply through lack of informal reading practice at the older age levels. These distinctions are ignored in all the studies reviewed. Thus McLeod (1965) set up as his criterion at least 1½ yr. of retardation; but since his children were above CA 10½ yr., many would presumably have been reading beyond RA 9 yr., that is, at a level adequate for many practical purposes. In fact, some of his duller Ss would have been reading at about their MA level; but McLeod (1965) published no details of the reading achievements of his Ss.

The question of whether retarded readers are also backward in the other basic subjects of spelling and arithmetic does not seem to have been investigated much since Burt's (1961) investigation of poor school achievers. He found that such children tended to be weak in all school subjects; but this was to be expected since his Ss were mostly duller than average. Also, of course, reading retardation would retard achievement in all other verbal subjects, and even in arithmetic in the later grades, when pupils are confronted with written problems.

METHOD

Subjects

Two groups of children, 54 retarded readers (experimental) and 54 adequate achievers (controls) were selected from primary schools in Sydney. Each group consisted of nine Ss from each of the six primary school grades, which range in CA from 6 yr. to 12 yr. To ensure that the groups were matched as closely as possible for the important variables of teaching method, curriculum, and socioeconomic class, both groups were drawn from the same classrooms in the same schools and at the same time of year. The schools were all situated in middle-class suburban areas to reduce the probability of selecting children whose poor achievement was due primarily to adverse conditions for home study. All Ss in both experimental and control groups were male.

All children were tested by the two authors. Retarded readers were first identified through class teachers. Each S was then individually screened on tests of visual and auditory acuity, and on Schonell's Graded Word Reading Test which involves only the untimed oral reading of single words. Children

^a Only one such case, an epileptic, has ever been found by the authors.

who fitted the criteria for reading retardation were individually administered 11 subtests of the WISC, (Mazes omitted), those failing to achieve a predetermined level of intelligence being excluded from the sample. (The criteria for reading retardation and WISC achievement are detailed in a later section). The selected sample was further given the Australian Council for Educational Research (A.C.E.R.) test of Mechanical Arithmetic (i.e., of arithmetical calculation) and a spelling test which were both individually supervised for each child.

Adequate achievers were first identified by rank in class at the last term examination. Where classes were divided into A and B streams on the basis of school achievement, the lower stream was chosen, since the retarded readers would also be there. Overall examination performance close to the median of the class was the criterion set for inclusion in the sample. Those selected in this way were screened first on the tests for sensory acuity, then the WISC, reading test, spelling test, and arithmetic test were all administered in the same way as for the experimental group.

Means for the controlled variables and for reading achievement are listed in Table 1.

Criteria for Reading and Intelligence

Failure to recognize single words, or to sound them phonetically, was the basic criterion of reading retardation, hence the choice of the oral reading test. The degree of retardation required for inclusion in the sample followed a sliding scale, that is, Grade 1: $\frac{1}{2}$ yr. retarded, Grade 2: 1 yr. retarded, Grade 3: $1\frac{1}{2}$ yr. retarded, Grade 4: 2 yr. retarded, Grade 5: $2\frac{1}{2}$ yr. retarded, and Grade 6: 3 yr. retarded. In fact, none of the Grade 1 children could read more than half a dozen two- or three-letter words; and none of the Grade 6 children was less than $3\frac{1}{2}$ yr. retarded in reading. The sample could therefore be described as a group of *very retarded* readers who manifested learning difficulties in the basic reading processes.

The criteria for selection on the basis of intelligence posed a problem. One of the aims of the study was to examine the amount of scatter between verbal and nonverbal subtests, while the a priori defini-

tion of reading retardation demanded children of at least average intelligence. Obviously, to insist upon an average Full Scale IQ would be to limit the amount of possible scatter, for large differences between Verbal (V) and Performance (P) Scales would result in below-average scores on the combined scale in some cases. As a compromise, it was decided to insist on an IQ of at least 90 on *either* the V or P Scale, but not necessarily on both combined, so that children with widely discrepant scores could be included in the sample. As it eventuated, 6 of the 54 experimentals and one of the controls, had Full Scale IQs below 90. The lowest of these was IQ 80.

RESULTS

The means of both samples were at about the middle of the average range of intelligence, the difference in mean Full Scale IQ being only 5.68 points, (Table 2). This small difference could hardly explain the great discrepancy in reading achievement between the groups. Nevertheless the difference proved to be statistically significant: for Full Scale IQ, $t = 3.21$, $p < .01$; for V IQ, $t = 3.34$, $p < .01$; and for P IQ, $t = 1.81$, $df = 106$, ns .

Differences between groups for all standard deviations shown in Table 2 were also significant beyond the .01 level.

Further investigation of WISC scatter consisted of an analysis of $V \sim P$ IQ differences, and also of analysis of individual subtest differences.

Differences between Verbal and Performance IQ

Distributions of V and P scale scores are listed in Table 3. Kruskal-Wallis one-way analyses of variance were carried out for the various pairs of distributions. For the $V > P$ distributions, the ranks for the experimentals were significantly greater than for the controls ($\chi^2 = 4.02$, $df = 1$, $p < .05$). For the $P > V$ distributions, ranks for the experimentals were also greater than for the controls ($\chi^2 = 4.55$, $df = 1$, $p < .05$). Within the experimental group, however, there was no significant difference between $V > P$ and $P > V$ ($\chi^2 = .75$, ns). In other words, the experimentals had greater $V \sim P$ discrepancies than the controls in both directions, but the experimentals as a group were about evenly distributed in either direction.

From Table 3 it appears that the $V \sim P$

TABLE 1

SAMPLING DATA FOR EXPERIMENTAL AND CONTROL GROUPS

Variable	Experimentals	Controls
CA (yr.)		
<i>M</i>	9.54	9.23
<i>SD</i>	1.98	1.88
Reading age (yr.)		
<i>M</i>	6.94	8.98
<i>SD</i>	.85	1.52
Socio ^a class		
<i>M</i>	3.41	3.59
<i>SD</i>	1.61	1.67

Note.—^a*N* = 54 in each group.

TABLE 2

IQ, MEANS, AND STANDARD DEVIATIONS FOR
EXPERIMENTAL AND CONTROL GROUPS

	Experimental	Control
Full scale		
<i>M</i>	100.85	106.53
<i>SD</i>	11.22	6.62
Verbal		
<i>M</i>	99.70	105.85
<i>SD</i>	11.29	7.38
Performance		
<i>M</i>	101.93	106.15
<i>SD</i>	14.54	8.77

index becomes discriminating at differences beyond about 15 points of IQ. However, further analysis shows this index to be an insensitive one, since the majority of retarded readers scatter below the test norms on subtests contained in *both* V and P scales. This would of course reduce the $V \sim P$ difference scores. The number of Ss who in the experimental group had at least one subtest score which was one standard deviation or more below the WISC subtest mean, (i.e., a score of 7 or below) was tallied. Of the 43 Ss who had at least one subtest score of 7 or less, 12 were entirely on V subtests, 11 were entirely on P subtests, and 20 were on both scales. This finding also disposes of any idea that there may have been two syndromes of reading retardation contained in the distributions, that is, one $V > P$ syndrome and one $P > V$ syndrome.

TABLE 3

FREQUENCY OF VERBAL IQ AND PERFORMANCE
IQ DIFFERENCES

$V \sim P$ Differences	$V > P$		$P > V$	
	Control	Experimental	Control	Experimental
1-4	10	2	5	9
5-9	6	9	8	3
10-14	7	6	6	7
15-19	2	4	3	2
20-24	2	2	1	2
25-29	0	1	0	2
30-34	0	0	1	3
35-39	0	1	0	0
Total	27	25	24	28

Note.—Zero differences: controls, 3; experimentals, 1. V = verbal IQ; P = performance IQ.

Comparison of Subtests

Means, standard deviations, and *t*-test results for the 11 WISC subtests are shown in Table 4.

From Table 4 it is apparent that the IQ difference between the two groups was due principally to differences in the Information, Arithmetic, Coding, and to a lesser extent, the Vocabulary and Digit Span subtests. The failure to find a consistent $V < P$ trend in the previous analysis was due mainly to the importance of the Coding subtest in differentiating between the two groups.

Both McLeod (1965) and Neville (1961) explained the consistently found low Information score as an *effect* of reading disabil-

TABLE 4
MEANS AND STANDARD DEVIATIONS OF WISC SUBTESTS

Subtest	Experimentals		Controls		
	<i>z</i>	<i>SD</i>	<i>z</i>	<i>SD</i>	
Information	9.74	2.74	11.59	1.79	4.19***
Comprehension	10.83	2.68	10.52	2.35	.64
Arithmetic	9.82	2.61	11.87	2.59	4.26***
Similarities	10.09	2.68	10.26	1.81	.39
Digit span	8.94	2.54	10.02	2.48	2.25*
Vocabulary	10.13	2.75	11.39	2.08	2.73**
Picture completion	11.00	3.49	11.61	2.88	.99
Picture arrangement	10.67	2.84	10.79	2.66	.25
Blocks	10.67	2.81	10.61	2.22	.01
Object assembly	9.39	2.43	10.04	2.07	1.45
Coding	9.50	2.93	11.30	2.15	3.64***

* $p < .05$, *df* = 106.

** $p < .01$, *df* = 106.

*** $p < .001$, *df* = 106.

ity rather than as a symptom. That is, they claimed that much of the information contained in the items may be gained through reading, so that the retarded reader is at a disadvantage on this subtest. However, if this were so, one would expect the discrepancy between the groups on this subtest to increase with school grade. That is to say, relevant information is unlikely to be gleaned from books in the first two grades of primary school (CA 6 yr. through to 7.9 yr.) when the children are in the early stages of learning to read; but as reading becomes fluent in the later grades, children might be expected increasingly to read books for interest and pleasure, thus gaining information relevant to the subtest. The same claim and the same argument might be applied to the Vocabulary subtest. However, Table 5 shows no indication of any trend toward increasing discrepancies either for Information or Vocabulary; the data do not warrant a test for significance of Grade \times Group interactions.

The difference between the groups on Coding could also be explained as an effect of reading retardation, since the task involves the rapid reading and writing of a code. Again, if this were true, the gap should in-

TABLE 5
MEAN SUBTEST SCORES FOR SCHOOL GRADES

School grade	Information		Vocabulary	
	Experimentals	Controls	Experimentals	Controls
1 + 2	9.89	11.22	10.11	11.61
3 + 4	9.83	11.94	10.67	11.73
5 + 6	9.45	11.33	9.61	10.83

Note.— $N = 18$ in each group.

crease with age, as the adequate achievers gain relatively more practice in reading and writing fluency. An analysis of variance was carried out to test whether the Groups \times Grades interaction was significant; but an F test was unnecessary since the interaction mean square was smaller than that of the remainder.

If it could be assumed that the retarded readers' group and the adequate achievers' group were representative samples, the findings in Table 3 could be taken as a valid indication of the differences between the groups. That the two groups should differ in Full Scale IQ is not surprising; since a downward scatter among certain of the subtests

TABLE 6
ADJUSTED MEANS, SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS,
AND CORRELATIONS OF SUBTESTS WITH FULL SCALE IQ

Group	Verbal subtest					
	Information	Comprehension	Arithmetic	Similarities	Digits	Vocabulary
Experimentals	10.17	11.18 ^R	10.26	10.46	9.14	10.49
Controls	11.10	10.17 ^R	11.47	9.89	9.82	11.03
\bar{x}	5.19*	4.92* ^R	7.52**	2.27	1.64	1.95
F	.579	.447	.499	.536	.254	.484
	Performance subtest					
	Picture Completion	Picture Arrangement	Blocks	Object Assembly	Coding	
Experimentals	11.29	11.18 ^R	11.09 ^R	9.80	9.76	
Controls	11.08	10.29 ^R	10.19 ^R	9.63	11.04	
\bar{x}	1.45	4.65* ^R	4.36* ^R	<1.00	6.72*	
F	.347	.583	.541	.585	.327	

Note.— R = reverse direction.

* $p < .05$.

** $p < .01$.

was postulated, an overall IQ difference was to have been expected.

However, it could be argued that the adequate achievers' group may have been subject to sampling error, since the mean IQ was 6 points above the mean of the general population. This would not be a particularly good argument, first because the WISC was not standardized on an Australian population, and second, because it does not follow that average class achievement corresponds exactly to average IQ. But it is certainly true that adequate achievement was a somewhat elastic concept, based as it was upon rank order in class, rather than on an absolute score. For this reason, covariance analyses, by which adjustments for mean IQ differences could be made, were carried out. The adjusted means and significance of differences are shown in Table 6.

As the assumption of homogeneous variances required for these tests was not met, a square-root transformation of the Full Scale IQ scores was undertaken, and this reduced the differences to acceptable limits. For all tests, the assumption of parallel regression lines was substantiated.

It will be seen from Table 6 that, when Full Scale IQ differences are taken into account, the controls are superior only on the Information, Arithmetic, and Coding subtests. The experimentals are shown to have been significantly better than expected in view of their lower IQ on the Comprehension, Block Design, and Picture Arrangement subtests.

Differences in Spelling

Since the A.C.E.R. Spelling Test, the only available one with Australian norms, has a lower limit at Grade 3 level, this test was supplemented by an ad hoc list of words

TABLE 7
ANALYSIS OF VARIANCE: SPELLING TEST

Source	SS	df	MS	F
Between groups (A)	6,611.34	1	6,611.34	137.74*
Between grades (B)	7,287.94	5	1,457.59	30.02*
A × B	254.82	5	50.97	1.05
Remainder	4,661.78	96	48.56	
Total	18,815.88	107		

* $p < .001$.

TABLE 8
ANALYSIS OF VARIANCE: ARITHMETICAL COMPUTATION

Source	SS	df	MS	F
Between groups (A)	564.90	1	564.90	24.59*
Between grades (B)	9,967.19	5	1,993.44	86.76*
A × B	216.38	5	43.28	1.88
Remainder	2,205.78	96	22.98	
Total	12,954.25	107		

* $p < .001$.

drawn from the reading primers of Grades 1 and 2. The analysis of variance shown in Table 7 was therefore carried out on the raw scores, that is, on the number of words correctly spelled. The test was administered individually, and there was no time limit.

The results shown in Table 7 show a highly significant difference in spelling achievement between the two groups in favor of the controls, and no Groups × Grades interaction effect.

Differences in Arithmetic

The A.C.E.R. test of Mechanical Arithmetic was administered individually to each child. This test is untimed, and consists solely of calculation sums. The reading of words is required for only one item, and this was read out to the Ss as necessary. The reading and writing of numbers was involved, but it was determined that all children could recognize numbers from 0 to 9. No child failed because of inability to read, but some of the experimental Ss failed on one or two items through a tendency to write number sequences in reverse, (e.g., 137 written as 731). A number of the younger experimental Ss either did not recognize arithmetical signs, or did not understand subtraction, multiplication, or division processes. For the older ones, errors occurred mainly in calculation.

Since the test norms have a lower limit at Grade 3 level, ad hoc items were added from elementary text books appropriate to Grades 1 and 2. Raw scores were used in the analysis of variance in Table 8.

The results in Table 8 show a highly significant difference between the computational achievement of retarded readers and adequate achievers, and no Grades × Groups interaction.

DISCUSSION

The inferiority of the experimentals on the subtests of Vocabulary, Digit Span, and, more importantly, of Information, Arithmetic, and Coding may be referred to the factor analytic findings of Witkin, Dyk, Faterson, Goodenough, and Karp (1962). Analyzing the performance of 10-yr.-old boys, they found that Information and Vocabulary loaded highly on a factor of Verbal Comprehension, whereas Arithmetic, Digit Span, and Coding loaded highly on a factor identified as Attention and Concentration (though it seems that this could also have been considered as a Memory factor).

The hypotheses of McLeod (1965) and Neville (1961) that lower performances on the Information and Vocabulary subtests of the WISC are effects of reading retardation rather than correlated symptoms were not confirmed in the present investigation. The relevant analyses were based on the assumption that if they were effects, the discrepancies between the groups should become greater with age, owing to the increasing reading fluency of the controls.

The hypothesis put forward by the present authors is that the WISC deficits are not effects, but symptoms correlated with reading disability, and primarily of a verbal nature. All of the significantly low subtests except Coding require verbalization; but it is possible that the experimentals' poorer performance on that subtest reflect a failure to use an effective labeling strategy as a memory aid, resulting in increased time taken in checking the code key for each symbol. Similarly it is possible that their poorer performance on Digits and Arithmetic was due to failure to use subvocal rehearsal as a strategy for keeping the items in mind. Whether these speculations are valid can of course be proved only by further experimentation; the reason they are put forward at all is that these same children were found in another, as yet unpublished, part of the investigation, to have had histories of delayed speech development and articulation difficulty in infancy significantly more often than the controls.

It is more difficult to account in these terms for the relatively good showing of the

experimentals on the Comprehension subtest. Despite the negative findings of Neville (1961), McLeod (1965), and Belmont and Birch (1966), there is support for this finding from the earlier studies of Burks and Bruce (1955), Coleman and Rasof (1963), Dockrell (1960), and Robeck (1962). Some light may be thrown on this problem by reference to the factor analytic study of Cohen (1959), who found two verbal factors, one of retained knowledge impressed by formal education, upon which Information and Arithmetic had high loadings, and one of informal verbal learning, upon which Comprehension loaded highly. It may also be noted that this subtest seems to require less verbal precision on the part of the testee than any of the other subtests.

The findings that retarded readers did relatively well on Block Designs and Picture Arrangement are supported by Neville (1961). Other investigators have found other P subtests to favor retarded readers: Picture Completion (McLeod, 1965) and Object Assembly (Belmont & Birch, 1966).

The finding that retarded readers were so much poorer in simple arithmetical calculation than adequate school achievers was somewhat surprising. Certainly this failure cannot be explained as effects of reading disability, for no reading was involved. The authors can think of nothing in terms of differential curricula or selection criteria which would have accounted for such a difference. In fact, a number of experimentals in Grades 4, 5, and 6 were repeating grades because of reading disability, and hence would have been exposed to more than 1 yr.'s learning of arithmetic at that grade level. On this basis one might have expected a superiority in Arithmetic. This is then another example of failure in formal learning, and again the hypothesis concerning verbal strategies may be tentatively invoked, since verbal labeling of numbers and processes is probably an important part of arithmetical calculation. Incidentally, this inferiority, as for reading, was present consistently from the earliest school grade.

That retarded readers were greatly inferior to controls in spelling is less surprising. In acquired brain injury, loss of function may be relatively specific; but in developmental

dysfunction, as reading retardation appears to be, it is to be expected that where there is an input disorder, there will be a corresponding output disorder; that is, where recognition (i.e., reading) does not take place, neither can recall (i.e., spelling) take place.

Retarded readers, it seems, can generally be regarded as retarded learners in all the basic school subjects at all primary grade levels, though of course there may be individual exceptions to this rule.

REFERENCES

- ALTUS, G. A WISC profile for retarded readers. *Journal of Consulting Psychology*, 1956, 20, 155-156.
- BELMONT, L., & BIRCH, H. G. The intellectual profile of retarded readers. *Perceptual and Motor Skills*, 1966, 22, 787-816.
- BURKS, H. F., & BRUCE, P. The characteristics of good and poor readers as disclosed by the WISC. *Journal of Educational Psychology*, 1955, 46, 483-493.
- BURT, C. *The backward child*. (5th ed.) London: University of London Press, 1961.
- COHEN, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology*, 1959, 23, 285-299.
- COLEMAN, J. C., & RASOF, B. Intellectual factors in learning disorders. *Perceptual and Motor Skills*, 1963, 16, 139-152.
- DOCKRELL, W. B. The use of WISC in the diagnosis of retarded readers. *Alberta Journal of Educational Research*, 1960, 6, 86-91.
- HIRST, C. S. The usefulness of a two-way analysis of WISC subtests in the diagnosis of remedial reading problems. *Journal of Experimental Education*, 1960, 29, 153-160.
- GRAHAM, E. E. W-B and WISC scattergrams of successful and unsuccessful readers. *Journal of Consulting Psychology*, 1952, 16, 268-271.
- MCLEOD, J. A comparison of WISC subtest scores of pre-adolescent successful and unsuccessful readers. *Australian Journal of Psychology*, 1965, 17, 220-228.
- MONEY, J. (Ed.) *Reading disability*. Baltimore: Johns Hopkins Press, 1962.
- NEVILLE, D. A comparison of the WISC patterns of male retarded and non-retarded readers. *Journal of Educational Research*, 1961, 54, 195-197.
- ROBECK, M. C. Children who show undue tension when reading; a group diagnosis. *International Reading Association Conference Proceedings*, 1962, 7, 133-138.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation: Studies of development*. New York: Wiley, 1962.

(Received July 6, 1967)

CONTRIBUTION OF SELECTIVE POSITIVE REINFORCEMENT AND THERAPEUTIC INSTRUCTIONS TO SYSTEMATIC DESENSITIZATION THERAPY¹

HAROLD LEITENBERG, W. STEWART AGRAS, DAVID H. BARLOW

AND DONALD C. OLIVEAU

University of Vermont

The contribution of therapeutically oriented instructions and verbal positive reinforcement to systematic desensitization therapy was investigated. Ss were 30 female college students with severe fear of snakes. Following treatment, the increase in approach behavior was significantly greater for Group 1 (graded hierarchy, relaxation, reinforcement, and therapeutic instructions) than for Group 2 (graded hierarchy and relaxation only). Group 2 did not differ significantly from a no-treatment control group. These results pose certain problems for reciprocal inhibition conceptions, and suggest that the effects of systematic desensitization cannot be attributed solely to the variables of a graded hierarchy plus relaxation.

According to Wolpe (1958) and Wolpe and Lazarus (1966) the effects of systematic desensitization therapy can be explained by two variables: (a) the temporal association of a state of relaxation with scenes that evoke anxiety and (b) the presentation of these scenes in a graded hierarchy. Relaxation is assumed to inhibit anxiety and to become conditioned to the formerly anxiety-evoking stimuli after a sufficient number of pairings. When the patient is asked to imagine himself in a series of situations that progress from the least to the most anxiety evoking, each succeeding item is presumed to elicit minimal amounts of anxiety which can be overcome by relaxation. Using severe fears of snakes and spiders as assumed prototypes of neurotic phobias, several investigators have confirmed the importance of combining relaxation with step-by-step exposure to the items in the hierarchy (Davison, 1968; Lomont & Edwards, 1967; Rachman, 1965).

The question remains, however, whether this combination alone can account for the effects of systematic desensitization therapy or whether there are other contributing factors. A review of the procedures used in systematic desensitization (Wolpe, 1958; Wolpe & Lazarus, 1966) suggests that therapeutically oriented instructions and positive re-

inforcement of approach behavior may also be important factors.

The usual therapeutic instructions indicate to the patient that a favorable outcome is likely and that gradual progress will occur during therapy. By themselves, such instructions may be relatively ineffective (cf. Lang, Lazovik, & Reynolds, 1965) but when added to the other basic ingredients of the systematic desensitization procedure they may increase the likelihood of a desirable outcome. In the area of hypnosis, Orne's research (1959) provides impressive evidence that Ss tend to fulfill explicit and implicit expectations in a situation.

In systematic desensitization, positive reinforcement in the form of selective attention and praise may be provided for reports of no anxiety, for completion of items in a hierarchy, and for reports of improvements outside the office. In an uncontrolled case report, Lazarus, Davison, and Polefka (1965) have previously argued that combining operant reinforcement procedures with desensitization procedures can facilitate the reduction of phobic avoidance behaviors. Truax (1966b) suggests that selective positive reinforcement is present in most forms of psychotherapy and may account for the similarity in outcomes often reported with different therapies. He has found (1966a) that response contingent reinforcement is present in Rogerian nondirective therapy.

¹This research was supported in part by a grant (MH-13651-01) from the National Institute of Mental Health, United States Public Health Service.

The purpose of the present experiment is to investigate whether instructions and verbal positive reinforcement are factors contributing to the success of desensitization therapy.

METHOD

Subjects

Thirty female *Ss* were selected from the undergraduate student body at the University of Vermont as follows: Questionnaires were distributed to the undergraduates at their weekly dormitory meetings. They were told that the Departments of Psychiatry and Psychology were looking for volunteers for an experiment investigating different methods of psychotherapy and the physiology of fear. They were then given a questionnaire asking them to rate their fear of nonpoisonous snakes on a 5-point scale. Following later evaluation of the questionnaire, all those students who noted that they were at least "definitely tense" at the sight of a nonpoisonous snake and who indicated a willingness to take part in the experiment were telephoned for an appointment. Since very few males met the criteria, the study was restricted to female undergraduates, ages 18-22.

Pretest

Before the pretest, *Ss* were interviewed to clarify their responses on the questionnaire and to gain some history about their fear of snakes. Possible *Ss* were then asked to enter a 15 × 10 ft. room and told:

At the end of the room there is a harmless green snake securely enclosed in a glass cage. The purpose of this session is simply to find out how afraid you are of the snake. To do this, I shall ask you to try to do various things with the snake. At no time will I force you to do anything which you are too afraid to do.

The glass cage was 6 × 10 × 8 in. and was covered by wire grating. The green snake (North American colubrid snake, *Opheodrys Aestivus*) was small and thin, measuring about 1 ft. in length. The floor of the room was marked off at 15-, 10-, 5-, 4-, 3-, 2-, and 1-ft. intervals from the cage. In order to insure some uniformity of activity, the snake was poked just before the assessment to provide some movement during the test. Approach to the snake was measured on a 14-point scale and the *S* qualified for the experiment by failing to reach Point 10, which consisted of reaching into the cage barehanded and touching the snake once. If the *S* qualified, she was matched with two other *Ss* who achieved similar scores and the three *Ss* comprised one block. Each *S* in this block was randomly assigned to one of three groups. The *Es* who administered the pre- and posttests did not know to which group the *S* was assigned. Similarly, the *Es* who performed the therapy were not aware of the pretreatment scores of the *Ss* on the scale. The *Ss* in Groups 1 and 2

were distributed equally between the two therapists. At the end of the pretest all *Ss* were told, "We are administering this test to a number of students and we will shortly decide which students will be selected for the experiment. The secretary will contact you." They were all asked not to discuss the experiment with other students.

Treatment

Systematic desensitization was administered to Groups 1 and 2 with different instructions and different reinforcement procedures. Group 1 (graded hierarchy, relaxation, therapeutic instructions, and positive reinforcement): *Ss* in this group were given therapeutically oriented instructions, the essentials of which were:

During these sessions we are going to be using a therapy that has been found effective in curing various fears, such as fears of snakes. In order to gain more understanding of how it works we are going to be recording physiological responses. . . . The way this therapy works is that by learning to relax deeply while imagining these scenes you will become less afraid of them. This will be true not only in imagination but also in the real situation.

After the initial instructions were completed and it appeared that *S* understood them, electrodes for monitoring galvanic skin response (GSR) were attached, and possible fears of being electrically shocked were allayed. GSR readings were taken throughout each entire session using a Grass Polygraph, model 7, with a direct current of 50 μ passing between silver, silver chloride palm, and wrist electrodes. In Session 1 *Ss* were also told that the button resting near their right thumb was to be pressed when they felt any anxiety. Pressing the button marked the recording paper.

In the first two sessions *Ss* were also trained in deep muscle relaxation using a 15-min. tape edited from a longer tape, originally made by Jacobson (1938). *Ss* were instructed to practice the relaxation procedure between these two sessions and during the second session the therapist inquired about the success and frequency of relaxation and answered any questions that *S* had. In the third session desensitization was begun.

A hierarchy of 27 items containing scenes involving snakes was used. A large number of items was employed to avoid big "gaps" in the hierarchy. This reduced the need for forming a subhierarchy or breaking down the item into more gradual steps for different *Ss* and kept the number of items that each *S* could be exposed to about the same.

The following is an example of a subhierarchy. Item 4 in the hierarchy consisted of a scene in which *S* imagined himself standing 10 ft. from a snake, and Item 5 consisted of a scene in which *S* imagined standing 5 ft. from a snake. If after 20 presentations *S* continued to report anxiety to Item 5, then a subhierarchy was employed consisting of

scenes in which *S* imagined himself standing 9, 8, 7, and 6 ft. from a snake.

Before introducing the first item, *Ss* in Group 1 were given the following instructions:

I am going to ask you to visualize various scenes involving the snake. I want you to report any anxiety or disturbance caused by a scene by pressing the button. This procedure has been found to be successful in reducing fears not only in the process of visualization, but also in the real situation.

This last statement was repeated until *E* was sure that *S* understood it.

Desensitization was begun with Item 1. Each item was read exactly as written from the work sheet and 10 sec. were allowed to elapse before *S* was instructed to "stop visualizing that and go on relaxing." Relaxation was continued for 30 sec. at which time the item was again presented. If *S* did not signal anxiety for 2 successive presentations of a particular item, *E* proceeded to the next item: "Now I want you to imagine a different scene." If, on the other hand, the *S* signaled anxiety, he was instructed to "stop imagining that and go on relaxing," and the same item was presented again. If anxiety was still aroused after 10 presentations, *E* dropped back to the previous item and presented that until it was imagined twice without anxiety, and then attempted the arousing item once again. The 30-sec. interval was kept constant between items. If, after 20 presentations of the same item either in 1 session or more, *S* still reported anxiety, a subhierarchy for that item was entered. All *Ss* who participated in the experiment were seen until either they completed the hierarchy or for 10 sessions (excluding the 2 sessions used for training in relaxation). Each session was half an hour in length.

Verbal praise such as "very good" or "you're doing fine" was given for any successful completion of an item, and at the end of the session progress was praised. Inquiries as to the clarity of visualization and depth of relaxation were frequently made.

At the end of the third desensitization session, *S* was given the following instructions.

As I told you at the beginning, this method has been useful in reducing fears. You have been progressing well in the visualization of the snake. I would like you to have an opportunity today to practice with the live animal. The snake is in a cage in the next room. The secretary will show you in. You may do as you like for not more than 15 minutes. Don't do anything that you find frightening; this is simply practice with the real thing.

The *E* inquired about success at the next session and any positive change was praised. The same instructions were repeated at the end of Sessions 6 and 9 if *S* had not completed the hierarchy by then.

Group 2 (graded hierarchy and relaxation only): *Ss* in this group were treated in the same manner with the following exceptions: (a) The instructions

were designed to eliminate the expectation that therapeutic benefit might be present in the treatment. (b) Positive reinforcement was *not* given for completion of an item on a hierarchy, nor was it given for reports of progress during the practice sessions. The essential instructions given Group 2 were the following:

You are taking part in an experiment on the visualization (or imagining) of snakes. We are going to record your physiological response on this machine while you visualize these objects. Because we want you to be able to visualize each scene very clearly we are going to teach you to relax during this session and the next. In this way there will be a minimum of extraneous factors in our recordings.

At the end of the third systematic desensitization session the rationale given for the practice session was: "I would like you to have an opportunity today to refresh your memory with the live animal in order to help you form a clear image."

Group 3 (no treatment): *Ss* in this group served as a control to determine the amount of improvement on the posttest in the absence of intervening treatment. After the other two *Ss* in each control *S*'s block finished the treatment, the control *S* was contacted for the first time since the pretest.

Posttest

The *Ss* in Groups 1 and 2 were assessed as soon as possible after treatment was completed. The median time was 3 days. The instructions given to all *Ss* were the same as those for the pretest. The *Es* administering the posttests were not aware of the conditions under which each *S* had been treated and the tests were conducted in such a way as to prevent the *E* from discovering in conversation prior to the test which *Ss* had been treated and which had not. One *S* from each group failed to complete the experiment. The data for these *Ss* were replaced according to Yates' formula for missing data (Cochran & Cox, 1957). The less appealing alternative was to discard the data from two complete blocks of matched *Ss*.

At the completion of the posttest, *E* asked questions designed to determine what group the *S* had been in. Twenty-six out of 27 *Ss* were correctly identified. One of the Group 2 *Ss* explained the purpose of the experiment in terms more suitable for a Group 1 *S*. Her data, nevertheless, were included in the analysis.

RESULTS

Since there was some question as to whether the data met the interval scale requirements of parametric statistical analyses, both parametric and nonparametric tests were carried out. In all instances both tests led to identical conclusions, so that only the values of the parametric tests will be reported.

Table 1 presents the mean scores the Ss received on tests of their ability to look at the live snake, put their hand close to the snake, touch the snake, etc. These tests were given before and after treatment and higher scores indicate greater approach behavior to the snake. For example, a score of 10 means the S touched the snake momentarily, whereas a score of 6 means that the S looked down at the live snake, but refused to put her gloved hand in the cage to the point where the knuckle would be below the top of the cage.

The Ss in the three groups were matched on the basis of their pretest scores and as can be seen the means on the pretest are very close together. The slight differences did not approach statistical significance ($F = .88$, $df = 2/18$, $p > .05$).

On the posttest nine Ss in Group 1 (graded hierarchy, relaxation, therapeutic instructions, and positive reinforcement), six Ss in Group 2 (graded hierarchy and relaxation), and only two Ss in Group 3, the control group, showed an increase in their approach scores. Difference scores were obtained for each member of each group by subtracting the pretest score from the posttest score. An analysis of variance for randomized blocks was then performed comparing the mean difference scores for the three groups. This analysis yielded a significant F ratio ($F = 6.96$, $df = 2/15$, $p < .01$). (Three degrees of freedom for the error mean square were lost because Yates' formula for computing missing data was employed.) A Duncan multiple-range test indicated that the mean difference score for Group

1 was significantly greater than the mean difference score for the control group ($p < .01$) as well as significantly greater than the mean difference score for Group 2 ($p < .05$). The mean difference score for Group 2, the group which received neither therapeutic instructions nor selective positive reinforcement, did not differ significantly from the mean difference score of the control group ($p > .05$).

Although the difference score for Group 2 did not differ significantly from the difference score of the control group, a within-Ss analysis indicated a significant increase from pre- to posttest scores for Group 2, $t(9) = 3.50$, $p < .005$, but no such significant increase was noted for the control group, $t(9) = .68$, $p > .05$.

Data collected during therapy sessions concerning Ss ability to progress through the hierarchy and her verbal and physiological reactions to items in the hierarchy are also of some interest. The mean number of hierarchy and branch items completed by Group 1 was 26.55 while the mean number of items completed by Group 2 was 22.88. A t test for matched samples indicated, however, that this difference was not statistically significant, $t(8) = 1.34$, $p > .05$. This trend is related to the finding that Ss in Group 1 had to have each item presented a fewer number of times than Ss in Group 2 before the criterion for going on to the next item was reached. (The criterion was two successive presentations without a report of anxiety.) The mean was 4.44 for Group 1 and 10.16 for Group 2, $t(8) = 1.15$, $p > .05$. Although neither of these differences are statistically significant, they are in the expected direction. Reinforcement for successful visualization of an item and for completion of an item may have increased the number of items completed and decreased the frequency of successive anxiety reports. The GSR data did not yield any significant difference between experimental groups, but it was quite clear that there was no one-to-one correspondence between verbal reports of anxiety and the presence of a GSR. Within each group a t test indicated that the percentage of item presentations which elicited an anxiety report was significantly greater than the percentage of item presentations which elicited a GSR ($p < .05$). This re-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PRETEST, POSTTEST, AND DIFFERENCE SCORES

Group	Pretest		Posttest		Difference between Pre- and Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Graded hierarchy, relaxation, therapeutic instructions, and positive reinforcement	6.40	2.17	8.61	2.35	2.21	2.00
Graded hierarchy and relaxation only	6.20	2.66	7.18	2.64	.98	.92
No treatment	6.60	1.96	6.80	1.75	.19	.92

sult is most likely due to adaptation of GSR activity following repeated presentations of an hierarchy item. Martin (1956) has previously reported a decline in skin conductance in successive free association segments of experimental interviews, and as Lacey (1959) points out "adaptation is a very common and striking phenomenon in all autonomic responses [p. 166]." Another possible interpretation is that physiological indexes of anxiety *extinguish* more rapidly than do verbal indexes of anxiety. (Also, verbal report may have a lower reaction threshold, that is, be more "sensitive," than GSR in cooperating Ss.)

DISCUSSION

The results of the present study suggest that therapeutically oriented instructions and selective positive reinforcement can enhance the beneficial effects of systematic desensitization therapy. The instructions given to Group 1 Ss indicated that the experiment would reduce their fears of snakes, while those given to Group 2 Ss indicated that the experiment was designed to learn about their physiological reactions while they imagined fearful stimuli. The therapist provided reinforcement by remarking "good," "excellent," "you're doing fine" when an S in Group 1 was able to visualize a scene without reporting anxiety, when an item in the hierarchy was completed, and when S reported progress in approaching the snake during practice sessions. Although both experimental groups showed a significant increase in their ability to approach the snake after treatment was completed, the magnitude of this increase was significantly greater for Group 1 (graded hierarchy, relaxation, reinforcement, and therapeutic instructions) than for Group 2 (graded hierarchy and relaxation). In fact, since Group 2 did not differ significantly from a control group which received only the pre- and posttests, it can be argued that either reinforcement or therapeutic instructions or their combination were major rather than minor contributors to the observed increase in approach behavior.

Although the combined absence of these variables significantly reduced the effectiveness of the systematic desensitization pro-

cedure, no attempt was made in this study to determine the *relative* effects of instructions and reinforcement in systematic desensitization. Presumably they are most effective when combined (cf. Ayllon & Azrin, 1964). This hypothesis is currently being tested in the authors' laboratory.

If selective reinforcement and therapeutically oriented instructions of the kind used for Group 1 are customarily a part of the systematic desensitization procedure, and examination of the detailed protocols provided by Wolpe and Lazarus (1966) suggest that they are, then the usual theoretical explanations of the effects of systematic desensitization are brought into question by these findings. Theoretical statements which stress almost exclusively the notion that relaxation physiologically inhibits anxiety and then becomes conditioned to the formerly anxiety-arousing stimuli would seem to be incomplete. Leitenberg and Agras (in press) have argued that relaxation facilitates the neutralization of anxiety-arousing stimuli but does not become conditioned to these stimuli. Following successful desensitization therapy, formerly anxiety-arousing stimuli seldom evoke a relaxed state—they simply no longer evoke anxiety. Relaxation may aid in the extinction of anxiety but the authors' findings suggest that therapeutic instructions and/or the step-by-step reinforcement procedure are largely responsible for the acquisition of adaptive behavior.

The results of the present experiment should not come as a surprise. In their own right both instructions and reinforcement have previously been shown to have significant behavioral effects. In psychotherapy and other therapylike situations Ss show changes in behavior in accordance with what they have been led to expect by past experience, instructions, and other placebo cues (e.g., Goldstein & Shipman, 1966; Orne, 1959). And operant conditioning has been employed to reduce the frequency of various abnormal behavior patterns (e.g., Ayllon & Haughton, 1962; Barlow, Agras, & Leitenberg, 1967; Ullman & Krasner, 1965).

A possibly important aspect of the systematic desensitization procedure, related to both instructions and reinforcement, is gen-

erally ignored in the literature. The precise instructions, an emphasis on behavioral measurement, and the structured design of systematic desensitization define for the patient exactly what behaviors are of interest and provide clear evidence throughout therapy that these behaviors are indeed changing in a gradual and systematic manner. The patient receives constant feedback from his own observations that the behavior in question can be changed in small steps, that he is being successful and is on the road to recovery. These changes are made possible by the small requirements of the graded desensitization procedure, and the importance of these changes is suggested to the patient by initial instructions, as well as by reinforcement during therapy when the therapist praises progressive change. Such *self-observed* signs of improvement may account for much of the success of all graded behavioral therapies, not only systematic desensitization. This question remains to be experimentally investigated.

REFERENCES

- AYLLON, T., & AZRIN, N. H. Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior*, 1964, 7, 327-331.
- AYLLON, T., & HAUGHTON, E. Control of the behavior of schizophrenics by food. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 343-352.
- BARLOW, D. H., AGRAS, W. S., & LEITENBERG, H. Control of classic neurotic "symptoms" through reinforcement and non-reinforcement. Paper presented at the First Annual Meeting of the Association for the Advancement of Behavior Therapy, Washington, D. C., September 1967.
- COCHRAN, W. G., & COX, G. M. *Experimental designs* (2nd ed.) New York: Wiley, 1957.
- DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, 73, 91-99.
- GOLDSTEIN, A. P., & SHIPMAN, W. G. Patient expectancies, symptom reduction and aspects of the initial psychotherapeutic interview. In A. P. Goldstein & S. J. Dean (Eds.), *The investigation of psychotherapy*, New York: Wiley, 1966.
- JACOBSEN, E. *Progressive relaxation*. Chicago: University of Chicago Press, 1938.
- LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington: American Psychological Association, 1959.
- LANG, P. J., LAZOVIEK, A. D., & REYNOLDS, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal Psychology*, 1965, 70, 395-402.
- LAZARUS, A. A., DAVISON, G. C., & POLEFKA, D. A. Classical and operant factors in the treatment of a school phobia. *Journal of Abnormal Psychology*, 1965, 70, 225-229.
- LEITENBERG, H., & AGRAS, W. S. Behavior therapy: Operant conditioning and related methods. In R. M. Jurjevich (Ed.), *Handbook of direct and behavior psychotherapies*. In press.
- LOMONT, J. F., & EDWARDS, J. E. The role of relaxation in systematic desensitization. *Behavior Research and Therapy*, 1967, 5, 11-25.
- MARTIN, B. Galvanic skin conductance as a function of successive interviews. *Journal of Clinical Psychology*, 1956, 12, 91-94.
- ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.
- RACHMAN, S. Studies in desensitization: I. The separate effects of relaxation and desensitization. *Behavior Research and Therapy*, 1965, 3, 245-251.
- TRUAX, C. B. Reinforcement and non-reinforcement in Rogerian Psychotherapy. *Journal of Abnormal Psychology*, 1966, 71, 1-9. (a)
- TRUAX, C. B. Some implications of behavior therapy for psychotherapy. *Journal of Counseling Psychology*, 1966, 13, 160-170. (b)
- ULLMAN, L. P., & KRASNER, L. (Eds.) *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.
- WOLPE, J., & LAZARUS, A. A. *Behavior therapy techniques: A guide to the treatment of neuroses*. London: Pergamon Press, 1966.

(Received August 11, 1967)

RELATIONSHIP OF THE A-B VARIABLE TO PATIENT SYMPTOMATOLOGY AND PSYCHOTHERAPY EXPECTANCIES¹

JURIS I. BERZINS

University of Kentucky

WILLIAM H. FRIEDMAN

Student Health Service, Indiana University

AND EDWARD SEIDMAN

University of Kentucky

In previous research, therapists' A-B status has predicted differential therapist effectiveness with schizoid versus neurotic patients, but patients' A-B status has not been studied. This study related the A-B scores of 60 male college clinic patients (As, controls, and Bs) to therapist ratings of presenting symptoms and patients' own complaints and role expectancies. As exhibited turning against the self, intropunitive symptoms, whereas Bs showed extrapunitive tendencies. Role expectancy data suggested, surprisingly, that As enter therapy expecting to unburden themselves in an active, productive manner, whereas Bs appear to seek rational guidance and correctives. With respect to these adjustive modes and role expectancies, therapist-patient complementarity, rather than similarity, may partially mediate the effectiveness of therapy interactions.

Recent research on therapist variables related to the outcome of psychotherapy points to the critical nature of the so-called A-B variable (see the review by Carson, 1967). On the basis of their scores on a small number of items taken from the Strong Vocational Interest Blank (the A-B scale), therapists categorized as "As" have been shown to achieve significantly better outcomes than "Bs" with schizophrenic patients (see reviews by Betz, 1962, 1967), whereas B therapists have performed better with neurotic outpatients (McNair, Callahan, & Lorr, 1962). Since the content of the A-B scale ostensibly involves varying degrees of interest in selected activities of a manual, technical, or mechanical nature (As dislike these activities, Bs like them), the A-B variable, despite its promise, is little understood.

Therapists' A-B status was initially reported to denote differences in "clinical style" in the treatment of schizophrenic patients in the Whitehorn-Betz studies (reviewed by Betz, 1967), where the As ap-

peared to behave in an active and experiential manner whereas the Bs tended to be passive or didactic in their conduct of therapy. Another correlate of A-B status appears to involve the field-dependence-independence dimension (Witkin, 1962): Bs were significantly more field independent and less variable than As in studies using actual therapists (Pollack & Kiev, 1963) and college students (Shows & Carson, 1965). The hypothesis that the A-B scale measures, beyond therapeutic "effectiveness," important dimensions of perceptual, cognitive, or communicative styles requires further examination.

The basic hypothesis implied by the original clinical studies, that there is an interaction between therapist type (A or B) and patient type (schizoid or neurotic), has received surprising support in several therapy analogue studies, using untrained male undergraduates as subject therapists (Carson, Harden, & Shows, 1964; Sandler, 1965). In these studies, the patient-type stimulus materials were usually adapted from Phillips and Rabinovitch's (1958) "avoidance of others" (AVOS) and "turning against the self" (TAS) symptom clusters as prototypic of schizoid and neurotic modes of adjustment, respectively. Results yielding significant Therapist \times Patient Type interaction effects, but

¹This study was partially supported by a summer fellowship to the first author from the University of Kentucky Research Foundation. The use of the facilities of the University of Kentucky Computing Center is gratefully acknowledged. The authors also wish to acknowledge the cooperation of the Indiana University Student Health Service staff.

opposite to the direction predicted by clinical studies, have been obtained when the amount of interaction between the participants has been attenuated, for example, by asking As and Bs to respond evaluatively to tape-recorded or written stimulus materials descriptive of the AVOS and TAS clusters (Carson & Klein, 1965; Kemp, 1966; Kemp & Sherman, 1965).

What variables mediate the apparent potency of $A \times AVOS$ and $B \times TAS$ pairings? Initial speculations have included reference to the possibility that sex role or social class related variables may be involved (Lorr & McNair, 1966). However, when Sandler (1965) asked male undergraduates to write out their typical reactions to stress, he found that the statements of As were significantly more often consistent with a trusting, intro-punitive, collaborative mode of adjustment, whereas Bs more often described themselves in suspicious, extrapunitive, avoidant terms. In other words, As and Bs resembled, in their respective modes of adjustment, those patients with whom they would be presumably less likely to be effective in actual therapy interactions.

Systematic differences in mode of stress adjustment, if reliably related to the A-B variable, should be highlighted within an actually disturbed population, that is, among patients rather than undergraduate Ss. The present study was designed to examine the relationship of patients A-B status to (a) therapist ratings of patient symptomatology evident in the first interview, (b) patients' own presenting complaints, and (c) patients' role expectancies regarding psychotherapy. The latter variable was selected to gain perspective on how the A or B patient construes his own and the therapist's probable behaviors prior to treatment.

Although this is the first time that the A-B scores of patients have been related to their own symptomatology and role expectancies, from Sandler's data, as previously cited, one would expect As to show symptoms consistent with a TAS (turning against the self, intro-punitive), and the Bs, with an AVOS (avoidance of others, extrapunitive) mode of adjustment. The role-expectancy measures

were expected to extend the known correlates of the A-B variable in a therapy-specific area.

METHOD

Patient and therapist samples. A battery of pre-therapy measures (see the following section) was filled out by 68 consecutive male patients prior to being assigned to eight therapists for initial interviews at the Student Health Service of Indiana University. The therapists comprised five psychiatrists, one psychologist, and two social workers. They interviewed from 2 to 16 patients each; four therapists accounted for 45 patients rated. No therapist was aware of the purpose to which his ratings were to be put.

Measures. The A-B scale version filled out by patients was a 21-item scale (comprising 15 items from the Strong Vocational Interest Blank and 6 items from the MMPI) resulting from a recent cross-validation² of the Kemp modification (Kemp, 1966) of the original Whitehorn-Betz scale. It was scored in the A direction (high scores yield A status, low scores yield B status). The A-B scale mean for the entire group was 12.02, with scores ranging from 5 to 19, suggesting that the lower (B) scale points may be somewhat underrepresented in this sample. Eight patients at the cutting points were randomly eliminated to yield two extreme groups (As and Bs) and a midmost group (ABs or controls) of 20 patients each. The As were defined as those Ss having scores between 14 and 19, Bs—scores between 5 and 9, and controls—scores between 10 and 13, inclusive.

The primary measure of presenting symptomatology was a therapist-rated checklist (15 items)³ obtained immediately following the initial interview. Each item was followed by a 5-point scale assessing the extent to which a given complaint was characteristic of the patient (1 equals not at all; 3 equals moderately; 5 equals extremely). Three ratings were designed as AVOS criteria (withdrawn, hallucinations or delusions, externalizes anger) and three as TAS criteria (depression, suicidal thoughts, internalizes anger). The remaining ratings (degree of disturbance, prognosis, likability, rigidity, etc.) functioned as fillers.

Patients' own complaints were assessed by a provisional checklist (32 items including fillers) with 5-point scales (1 equals not at all; 3 equals moderately; 5 equals extremely) asking patients how much a given problem "distressed, bothered, or worried you during the past week, including today." The items of this checklist were loosely adapted from the Psychiatric Outpatient Mood Scales (McNair & Lorr, 1964) as well as the Mooney checklist (Mooney & Gordon, 1950) and included items related to the TAS and AVOS symptom clusters. Sample TAS items are: depressed, dejected, feel worthless, lack

² Robert C. Carson, personal communication, March 1967.

³ All measures used in this study may be obtained from the senior author upon request.

of interest; AVOS: awkward in meeting people, slow to get acquainted, few or no close friends.

Patients' expectancies regarding their first hour of psychotherapy were assessed by a revised version of a 27-item Psychotherapy Expectancy Scale or PEI, described by Berzins (1966). The provisional version of this scale used here (60 items, including fillers) asks patients to indicate how strongly they expect a given patient or therapist behavior to take place during the next (here, the first) hour of treatment (1 equals not at all; 4 equals moderately; 7 equals very strongly). While ordinarily this scale is scored for four patient-role categories (Nurturant-Coercive, Critic, Self-Reliant, Cooperative), the validity of the scale for male patients is questionable at this stage, and therefore items are simply used as indications of self-reported expectancies among patients varying in A-B status.

RESULTS

Patients' A-B Status and Symptoms

The relation of patients' A-B status to therapists' ratings in the TAS and AVOS categories is summarized in Table 1. It is evident that patients scoring in the A direction on the A-B scale are rated higher than ABs (controls) and Bs on the TAS criterion items of depression, suicidal ideation, and internalization of anger. This finding clearly supports Sandler's results on the stress reactions of undergraduates.

Patients scoring in the B direction on the A-B scale, however, are not classifiable as showing AVOS symptoms, excepting their ostensibly greater (than for As) tendency to externalize anger. It is apparent that the relation of A-B status to internalization-exter-

nalization of anger is not linear: the middle group (ABs) is rated as least internalizing (most externalizing) of the three groups; the regression for internalization of anger departs significantly from linearity ($F_{lin} = 5.99$, $p < .01$), with a correlation ratio (η) of .40.

Turning to the patients' own presenting complaints, as reflected in their checklist responses prior to the first interview, Table 2 reports the items on which Pearson r 's significant at or beyond .10 were found.

The majority of complaints associated with A-B status confirm therapists' criterion ratings: a status appears related to feelings of depression and worthlessness, tearfulness, apathy, a tendency to sleep instead of studying, and thoughts which cannot be confided. These symptoms are consistent with the TAS symptom cluster. The one item related to scoring in the B direction on the A-B scale (have keen sense of moral obligation) does not conform to the AVOS category, nor is its meaning clear. This item is *negatively* related to therapists' ratings for hallucinations and delusions ($r = -.34$, $p < .01$) and positively related to good prognosis ($r = +.27$, $p < .05$), suggesting that this "complaint" may have an adaptive meaning.

The lack of association between B status and the AVOS symptom cluster may partially result from a heterogeneity of referents for AVOS-like symptoms in the measures used. Extremely low means for therapist ratings of hallucinations and delusions (Table 1), for

TABLE 1
RELATIONSHIP BETWEEN PATIENTS' A-B STATUS AND THERAPIST-RATED SYMPTOMS

Therapist ratings	Correlation with patients' A-B score	M			
		As	ABs	Bs	F
TAS criteria					
Depression	+.41***	3.30	2.74	2.30	6.38***
Suicidal ideation	+.29**	2.40	1.89	1.60	3.57**
Internalizes anger	+.27** (.40*)	4.05	3.16	3.65	5.44***
AVOS criteria					
Withdrawn	-.12	2.35	2.21	2.65	<1.00
Hallucinations, delusions	-.10	1.10	1.21	1.25	<1.00
Externalizes anger	-.28**	1.45	2.05	1.95	2.54*

Note.— p values here and elsewhere are for two-tailed tests. One patient in the AB group was not rated, therefore $N = 59$.

* This is a correlation ratio (η), calculated because of significant ($p < .01$) nonlinearity of regression. All other correlations are Pearson r 's.

** $p < .10$.

*** $p < .05$.

*** $p < .01$.

TABLE 2
RELATIONSHIP BETWEEN PATIENTS' A-B STATUS AND SELF-REPORTED SYMPTOMS

Abbreviated content	Correlation with patients' A-B score	M			
		As	ABs	Bs	F
Feel depressed	+ .38***	4.05	3.95	3.15	5.73***
Feel worthless	+ .29**	3.70	3.30	2.63	3.83**
Often feel like crying	+ .28**	2.75	2.74	1.95	2.64*
Lack of interest	+ .26**	3.30	2.90	2.35	1.92
Sleep not study	+ .24*	3.40	2.75	2.53	1.93
Can't confide thoughts	+ .24*	2.95	2.15	2.05	2.82*
Have keen sense of moral obligation	- .27**	2.88	3.20	3.50	<1.00

Note.—Due to omissions, $N = 19$ for three means, and $N = 17$ for one.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

example, suggest that the incidence of psychoticlike symptoms was low in this sample of college clinic patients. On the other hand, patients' self-reports in the AVOS category addressed themselves to more "benign" forms of AVOS-like problems (e.g., "awkward with other people," "few or no close friends," "slow to get acquainted," etc.). In almost all cases involving AVOS-type complaints, B patients exceeded ABs and As, although the means did not reliably differentiate the groups. Therapists and patients do not appear to have addressed themselves to the same issues, as indicated by the differential validities (correlations between patients' complaints and criterion ratings) of the TAS and AVOS categories. The patient-therapist correspondences on TAS symptoms were satisfactory (r 's ranged from +.24 to +.47, median +.34) while the coefficients for AVOS symptoms were consistently lower (r 's ranged from -.07 to +.32, median +.16).

In a number of instances, the association between A-B status and patients' complaints significantly departed from linearity ($p < .05$). The roughly U-shaped functions call attention to the possibility that the middle of the A-B score distribution (ABs) may constitute a meaningful group in its own right, but the characteristics of ABs are not easily interpreted. Thus, while on some self-report items which dealt with TAS contents (Table 2), ABs were intermediate between As and Bs, it was seen (Table 1) that therapists rated ABs as *least* internalizing of anger among the three groups. Curvilinear relation-

ships between A-B status and patients' complaints included the following: relative to As and Bs, ABs reported themselves as less easily influenced by other people, less influenced by others' moods, less irritated by others, less prone toward escapism, and less unsure of others' reactions, even though they complained more of insomnia (correlation ratios ranged from .25 to .41 for these items). But what at first might appear to be the "better adjustment" of ABs was not supported by two therapist ratings which designated ABs as having poorer prognoses and being less trusting than As and Bs (etas of .37 and .34, respectively).

The expected association between patients' A-B status and the TAS mode of adjustment appears to be generally supported in this sample, while the findings with regard to B and AB patients are unclear.

Patients' A-B Status and Psychotherapy Expectancies

The correlations between patients' A-B scores and their expectancies regarding therapy are given in Table 3, along with the means for the trichotomized sample. While no hypotheses were made with respect to expectancies, the findings regarding symptoms would suggest that As would display little energy in therapy interviews, if the clinical association of depression and anergy be taken as a guideline. That this is not the case is evident in the upper portion of Table 3. Patients scoring in the A direction on the A-B scale appear to see the therapist as basically

TABLE 3

RELATIONSHIP BETWEEN PATIENTS' A-B STATUS AND THEIR EXPECTANCIES REGARDING PSYCHOTHERAPY

Patient expectancy (How strongly do you expect...)	Correlation ^a with A-B score	<i>M</i>			
		<i>A</i> _s	<i>AB</i> _s	<i>B</i> _s	<i>F</i>
your therapist to mostly listen to you talk out your problems?	+ .40***	5.45	4.83	4.30	3.73**
to feel like opening up without any help from your therapist?	+ .37***	5.26	4.83	4.00	3.62**
to be the one who begins the talking?	+ .29**	4.50	4.00	3.75	1.03
to feel that it's all right to talk about your feelings toward your therapist?	+ .26**	5.80	5.50	5.00	1.73
to "carry the ball" conversationally?	+ .26**	4.85	4.05	4.30	1.52
to initiate the conversation?	+ .24*	4.45	4.28	3.75	1.21
to lead the way in bringing up topics to talk about?	+ .22*	4.60	4.39	3.75	1.52
your therapist to level with you?	- .33**	5.45	6.45	6.55	3.93**
your therapist to clearly announce his value judgments about your behavior?	- .22*	4.05	5.33	4.85	2.84*
your therapist to find out what's the cause of your problems?	- .22*	4.80	5.39	5.50	1.42

Note.—It was belatedly discovered that two patients in the AB group did not complete this measure, therefore $N = 58$.
^a All correlations are Pearson r 's since no significant departures from linearity of regression were observed.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

a good listener. In complement, they anticipate exerting considerable initiative in beginning and maintaining the flow of conversation, and they expect to talk about their feelings. This emphasis appears to be changed for B patients (lower portion of Table 3). While, as with symptoms, few items are related to A-B scores in the B direction, B patients appear to expect the therapist to "level with them," to make evaluative comments, and to be concerned with the etiology of the patient's problems.

DISCUSSION

Individual differences on the A-B variable appear to have some relation to the presenting complaints and therapy-related expectancies of college clinic patients. The results supported the expected association between A status and the TAS mode of adjustment, and A patients also appeared to expect themselves to play verbally active and productive roles in treatment. With the exception of a tendency of B patients to externalize anger, the expected relationship between B status and the AVOS mode of adjustment was not clarified by this study. The B patients' expectancies suggested that they anticipated encountering and dealing with a straightforward, analytical, teacherlike figure.

The role-expectancy data appear to be somewhat consistent with the A-B differences in "clinical style" reported by Betz (1967) as characteristic of A and B therapists. Betz pointed out that B therapists tended to be either passive or instructional in their interactions with schizophrenic patients; in the present study, B patients appeared to envisage a somewhat "didactic" interaction. The B patients' expectancies regarding "structure" in psychotherapy are also congruent with the presumably high level of cognitive differentiation (greater field independence) of B therapists (Pollack & Kiev, 1963) and undergraduates (Shows & Carson, 1966). The A-type therapists, according to Betz, were actively and experientially involved with their patients; here, A patients expected themselves to "carry the ball" conversationally and did not appear to expect much activity from the therapist.

The unexpected aspect of the present results regarding the A patient was that, despite presumably high levels of depressive suffering, the A patient anticipated being quite vocal about his troubles! Speculatively, this tendency toward active, productive communication, which seems incompatible with the clinical picture of depression, bespeaks a readiness to turn to other people in times of

distress. If Bs evidence such tendencies to a lesser extent than As, the failure to obtain a full replication of Sandler's data on undergraduates might partially be due to variables which themselves are a function of A-B status, for example, an interpersonally avoidant mode of adjustment to stress. If these speculations have merit, then one would expect more As than Bs to come for psychotherapy (it will be recalled that no S scored within the lower five "B" points on the A-B scale), whereas one would not necessarily expect any differences in rates of volunteering for psychological experiments (e.g., in Sandler's study). Since it is generally acknowledged that neurotic patients often refer themselves for treatment, while schizoid or schizophrenic patients seldom do so, one might hypothesize that the A-B dimension, when interacting with levels of subjective distress, predicts differential rates of turning to other people for help, with As exceeding Bs in this regard.

The actively approaching treatment styles of A therapists, reported by Betz (1967), thus would appear similar to the tendencies toward interpersonal approach inferred from A patients' role expectancies. If such continuities are genuine, then a most important hypothesis, raised by even a partial replication of Sandler's results, is that therapist-patient *complementarity*, rather than similarity, on the A-B variable may mediate the "effectiveness" results obtained in prior research. Swensen (1967) has also suggested from a differing point of departure that the A-B dimension may in part be an interpersonal approach-avoidance dimension: A therapists' approaching behaviors would mesh well with the needs of withdrawn schizoid patients, whereas the neurotic (TAS) patient may not require an especially approaching therapist. If A-type therapists, like A patients, show tendencies to turn against themselves under stress, the finding that A therapists perform less well with neurotic patients (McNair, Callahan, & Lorr, 1962) may be due to A therapists' "blind spots" which interfere with therapeutic communication. Effective interaction would again be indexed by complementarity rather than similarity. The argument with respect to B-type therapists and patients is more tenuous but bears further

investigation. The extent to which personal therapy and professional training mitigate against such blind spots is of course an empirical question.

Briefly put, the complementarity hypothesis leads to the prediction that A therapists would perform better with B patients, and B therapists—with A patients (than As with As and Bs with Bs), on both the approach-avoidance and blind spot bases. The improvements in outcome resulting from such matching may of course be slight. The data of the present study indicated only moderate relationships between A-B status and the variables found associated with it. Also, the measures used were not behavioral and the AVOS category in particular may have been inadequately measured. Some of the curvilinear relationships observed as a function of including the AB (middle) group, which has few precedents in work with this scale, suggests, however, that it would be desirable to include groups of AB therapists and patients in future studies.

While the present study, thus, can only point toward promising further work, it should be pointed out that there is continuing evidence regarding intriguing continuities between diverse samples (i.e., practicing therapists, college students, and, here, patients) with respect to psychological variables associated with A-B status.

REFERENCES

- BERZINS, J. I. The behavioral validation of a psychotherapy patient role-expectancy inventory. Unpublished doctoral dissertation, University of Connecticut, 1966.
- BETZ, B. J. Experience in research in psychotherapy with schizophrenic patients. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy*, Vol. 2. Washington, D. C.: American Psychological Association, 1962.
- BETZ, B. J. Studies of the therapist's role in the treatment of the schizophrenic patient. *American Journal of Psychiatry*, 1967, 123, 963-971.
- CARSON, R. C. A and B therapist "types": A possible critical variable in psychotherapy. *Journal of Nervous and Mental Disease*, 1967, 144, 47-54.
- CARSON, R. C., HARDEN, J. A., & SHOWS, W. D. A-B distinction and behavior in quasi-therapeutic situations. *Journal of Consulting Psychology*, 1964, 28, 426-433.
- CARSON, R. C., & KLEIN, S. J. The A-B therapist "type" distinction and perception of patient characteristics. Paper presented at the meeting of the

- Southeastern Psychological Association, Atlanta, 1965.
- KEMP, D. E. Correlates of the Whitehorn-Betz AB scale in a quasitherapeutic situation. *Journal of Consulting Psychology*, 1966, 30, 509-516.
- KEMP, D. E., & SHERMAN, D. Relationships between medical students' A-B Scale scores and their evaluation of schizoid and neurotic patients. Paper presented at the Southwestern Psychological Association, Oklahoma City, 1965.
- LORR, M., & MCNAIR, D. M. Methods relating to evaluation of therapeutic outcome. In L. A. Gottschalk & A. H. Auerbach (Eds.), *Methods of research in psychotherapy*. New York: Appleton-Century-Crofts, 1966.
- MCNAIR, D. M., CALLAHAN, D. M., & LORR, M. Therapist "type" and patient response to psychotherapy. *Journal of Consulting Psychology*, 1962, 26, 425-429.
- MCNAIR, D. M., & LORR, M. An analysis of mood in neurotics. *Journal of Abnormal and Social Psychology*, 1964, 69, 620-627.
- MOONEY, R. L., & GORDON, L. V. *Mooney problem check lists. (College Form)* Psychological Corporation, 1950.
- PHILLIPS, L., & RABINOVITCH, M. Social role and patterns of symptomatic behavior. *Journal of Abnormal and Social Psychology*, 1958, 57, 181-186.
- POLLACK, I. W., & KIEV, A. Spatial orientation and psychotherapy: An experimental study of perception. *Journal of Nervous and Mental Disease*, 1963, 137, 93-97.
- SANDLER, D. *Investigation of a scale of therapeutic effectiveness: Trust and suspicion in an experimentally induced situation*. (Doctoral dissertation, Duke University) Ann Arbor, Mich: University Microfilms, 1965. No. 66-1382.
- SHOWS, W. D., & CARSON, R. C. The A-B therapist "type" distinction and spatial orientation: Replication and extension. *Journal of Nervous and Mental Disease*, 1965, 141, 456-462.
- SWENSEN, C. H. Psychotherapy as a special case of dyadic interaction: Some suggestions for theory and research. *Psychotherapy*, 1967, 4, 7-13.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation: Studies of development*. New York: Wiley, 1962.

(Received August 21, 1967)

INCONSISTENT COMMUNICATIONS AND PSYCHOPATHOLOGY¹

NANCY G. BEAKEL AND ALBERT MEHRABIAN

University of California, Los Angeles

The present study investigated the presence of possible incongruity between verbal and nonverbal components of parental messages to a disturbed child. Video-taped family interactions of two groups of families in which there were two degrees of psychopathology of an adolescent member were employed. There were two different measures of incongruity between attitude conveyed in verbalizations and posture. The findings did not support the hypothesized relationship between incongruity and degree of pathology. The results did however indicate a relationship between degree of negative attitude conveyed by parental messages and rated psychopathology of the adolescent in the family. It was found that parents of the more disturbed adolescents verbally communicate more negative attitudes toward the adolescents than parents of the less disturbed adolescents.

The present study is concerned with incongruity between the verbal and postural attitude communications of parents of disturbed children. A number of investigators have suggested that inconsistent attitude communications of parents contribute to the psychopathology of their children. While initial theorizing along these lines, as exemplified by double-bind theory, related inconsistent attitude communications of parents to schizophrenia in their offspring (e.g., Bateson, Jackson, Haley, & Weakland, 1956; Weakland, 1961), investigators subsequently suggested that there is a correlation between the frequency of occurrence of inconsistent attitude communications in a family and the degree of psychopathology of the most disturbed member of that family (e.g., Haley, 1963). In the present context, "communication of attitude" refers to the degree of liking, preference, or positive evaluation expressed by one individual toward another. In addition to verbalizations (i.e., contents and tone of voice), facial expression, body position, or gesture can, for example, also convey attitudes. Thus, differences between attitudes conveyed simultaneously in differ-

ent components of a communication can yield incongruous attitude communications.

Assessment of the effects of consistent and inconsistent attitudes in content and tone of voice was studied by Mehrabian and Wiener (1967). Three degrees of attitude (i.e., positive, neutral, and negative) in the tone of voice were each combined with three degrees of attitude in content (i.e., meanings of words). They found that when vocal communication of attitude is inconsistent with verbal communication of attitude, normal addressees respond to the two-channel communication by subordinating the verbal content to the vocal component. If, for example, the word "dear" is said in a tone of voice which is independently judged as communicating a negative attitude toward the addressee, the consensual interpretation of the total communication is negative, at least for normal Ss who are addressees of inconsistent communications.

The decoding of inconsistent communications was also investigated by Mehrabian and Ferris (1967). Their Ss rated all possible combinations of positive, neutral, and negative attitudes conveyed facially and vocally. Inferred attitude in this case was found to be a linear function of the independent effects of the two components, with the facial component receiving $3/2$ the weight received by the vocal component. Thus, the findings of the Mehrabian and Wiener and Mehrabian and Ferris studies suggest that normal ad-

¹This research was supported by National Institute of Mental Health grants MH 08744 and MH 13509. Special thanks are due Elliot Rodnick, Michael J. Goldstein, Edward Gould, and Sigrid McPherson for their ratings of adolescents' psychopathology. The cooperation of Daphne Bugenthal in providing anchor audio and video-tape recordings is greatly appreciated.

addressees resolve inconsistency conveyed by various components of an attitudinal message by taking a weighted sum of the components. Both studies, therefore, do not provide support for the double-bind theory assumption that inconsistent attitude communications cannot be resolved (i.e., are confusing for the addressee).

The present study deals with a second aspect of the double-bind theory, namely, the assumption that inconsistent communications of attitude are more frequent in the families of children who exhibit greater degrees of psychopathology. Thus, the present study was designed to explore the incongruity between attitude communicated in verbalization and attitude communicated posturally for two groups of parents who differed in the degree of psychopathology manifested by their adolescent child.

The communication of attitude via posture requires some consideration prior to the presentation of the study. Mehrabian (1968) hypothesized that a more open position, greater relaxation, and a more forward orientation of the body toward an addressee communicates a more positive attitude. He found that a more forward orientation and greater relaxation communicate a more positive attitude to an addressee. In addition he found that an open position conveys a positive attitude only when the communicator is female. His results can be used to measure either the actual degree of positive-negative attitude communicated by the posture of a seated communicator, or to obtain an estimate of the kind of posture that an addressee expects provided he only hears the verbalization accompanying that posture. The latter, anticipated posture, when compared to the actual posture, provides one measure of incongruity in communication.

In the following study, attitudinal messages were taken from two groups of parents who differed in the degree of psychopathology manifested in their adolescent child. Independent judgments of attitude in verbalization and posture were obtained, on the basis of which measures of incongruity of attitude communication were computed. It was hypothesized that parents of the more disturbed adolescents exhibit a greater de-

gree of incongruity between their verbal and postural attitude communications to their children than parents of the less disturbed adolescents.

METHOD

Subjects. Eight Ss, four males and four females in the 20-30 age range were selected from volunteers.

Materials. Verbalizations and postures were taken from among audio- and video-tape recordings of adolescents and their parents. The families were part of a sample of 21 adolescents and their parents who discussed a family problem stemming from the child's disturbance. In no case were any of the participants aware of the recordings which were made through a one-way mirror. All family members were seated during the recording.

Three clinical psychologists who were familiar with the individual problems represented in the entire sample ranked the adolescents concerned according to severity of pathology. The clinicians were requested to rank order the 21 adolescents from most severe to least severe without regard to diagnostic classification. Kendall's Coefficient of Concordance (Winer, 1962) indicated satisfactory agreement among the three judges ($W = .85$). Rank scores were averaged over the three judgments in order to select pairs of parents to serve in the experimental groups. Parents of the 5 adolescents receiving the lowest average rank scores were assigned to the more pathological group, while parents of the 5 receiving the highest average rank were assigned to the less pathological group.

For each parent of both pathology groups, four to six statements averaging approximately 35 words in length were selected from the audio portion of the family discussion tapes to be used as verbalizations. In selecting these audio communications, the judge was presented only with the audio tapes and was instructed to select statements which satisfied the following criteria: (a) statements should be complete sentences, (b) statements should involve the parent speaker addressing the child directly, either by name or by use of the pronoun "you," and (c) the statements should concern some aspect of the child's behavior problem or personality. Due to the selectivity of these criteria, the number of statements obtained for each parent ranged from four to six. It should be noted that in selecting these verbal communications, the judge had no access to the nonverbal correlates of these communications and therefore could not have been influenced by the degree of consistency or inconsistency of the total communications.

In the subsequent part of the selection of materials, the judge identified the video portions corresponding to the already selected verbal communications. In those instances where there were more than four verbal communications for a parent, the judge selected the four communications in which the parent was maximally visible, and if there still were extras, randomly eliminated one or two of the

set to obtain only four video communications. In sum, these procedures yielded a final set of four audio-video communications for each parent.

The total set of 80 communications consisted of four combined audio-video messages of each of 10 mothers and 10 fathers. These communications were coded and randomized.

Procedure. Judgments of attitude conveyed in verbalizations were obtained from four Ss who used a 7-point positive-negative attitude scale. The Ss could indicate the overall attitude expressed in each verbalization on a scale ranging from "extreme negative," 1, to "extreme positive," 7. Before presentation of the verbalizations, Ss were given four anchor stimuli and the use of the scale was explained.

The same four Ss subsequently rated video segments on the same scale to provide "global judgments" of attitude conveyed by posture. A practice session was given using four sample video-taped postures as anchors for the scale. In both practice and rating sessions, only that portion of the monitor screen showing the body of the parent speaker was exposed to the Ss. Other members of the family and the head of the parent speaker were masked by a black shield placed over the screen. The position of the child receiver relative to the parent speaker was indicated by a white "X" on the shield. This shielding device restricted judgments of attitude conveyed by posture to positions of the body only.

Four additional Ss provided "weighted judgments" of attitude conveyed by a posture anticipated for a given verbalization. These Ss listened to the taped verbalizations. Their instructions were to imagine the posture assumed by the parent speaker of each verbalization in a situation in which the parent and the child are seated. For each verbalization, Ss described the posture they imagined by selecting the appropriate descriptive word from each of the following four pairs: open-closed, tense-relaxed, forward-back, and toward-away from. "Open" was defined as a position of the limbs allowing access to the speaker's body and "closed" as a position denying access. "Forward" was defined as a forward inclination of the torso and "back" as any other position, including erect positions of the body. "Toward" was defined as a position of the body in which the speaker's chest is oriented toward the child receiver and "away from" as a position in which the speaker's chest is turned away from the child. "Tense" and "relaxed" were considered self-explanatory and were not specifically defined. A practice session was included in which drawings were used to explain the four pairs of body positions.

This second set of four Ss subsequently viewed and provided weighted judgments of the video segments. As in the case of the global judgments of posture, the monitor screen was shielded, affording Ss a view of the parent speaker's body only.

RESULTS

The following set of measures were obtained for each of the 80 communications.

Judgment of attitude in verbalization: This was the rating of the degree of positive-negative attitude communicated by the verbal component of each communication, as rated on a 7-point scale.

Global judgment of attitude in posture: This was the rating of the degree of positive-negative attitude communicated by the postural component of each communication, as rated on a 7-point scale.

Weighted judgment of attitude in actual posture: This was the rating of the degree of positive-negative attitude communicated by actual posture, using a set of weightings for postural components taken from Mehrabian's (1968) study.

Weighted judgment of attitude in anticipated posture: This was the rating of the degree of positive-negative attitude in the posture which was imagined for a verbalization. These scores were computed on the basis of the weightings employed in the preceding measure.

Besides these four primary measures, two additional composite measures of incongruity were also computed for each of the 80 communications.

Global incongruity score: To obtain a measure of incongruity between judgments of attitude in verbalization and global judgments of attitude in posture, the absolute difference between the two attitude scores was used. These absolute difference scores ranged from zero to six.

Weighted incongruity score: A second measure of incongruity employed the weighted judgment of attitude from posture. This was the absolute difference between the weighted judgment of attitude in the actual posture accompanying a verbalization and of the attitude in posture anticipated for that verbalization.

In the case of each of the above six measures, the scores obtained for a given communication from each of four raters were averaged.

Analysis of variance for reliability coefficients (Winer, 1962) was computed for the four primary measures. For the judgments of attitude in verbalizations, $r_4 = .78$; for global judgments of attitude in posture, $r_4 = .47$. Weighted judgments of attitude in actual

posture yielded $r_4 = .63$, and weighted judgments of attitude in anticipated posture yielded $r_4 = .48$.

Computation of a correlation matrix for all dependent variables yielded a significant positive correlation between global judgments of attitude in verbalizations and weighted judgments of attitude in anticipated posture ($r = .36$, $p < .01$). A significant positive correlation was also found between global and weighted judgments of attitude in actual posture ($r = .40$, $p < .001$).

A 2 (Pathology) \times 2 (Parent Sex) \times 5 (Individuals) \times 4 (Communications) analysis of variance was performed on each of the following four dependent measures. In this design, individuals are nested under the pathology and sex conditions, and communications are nested under pathology, sex, and individuals conditions.

Analysis of variance of global incongruity scores indicates only one significant effect due to sex of parent ($F = 5.21$, $df = 1/60$, $MS_e = .28$, $p < .05$). The mothers are more incongruent (1.24) than the fathers (.98).

Analysis of variance of weighted incongruity scores indicates significance due to pathology group ($F = 4.12$, $df = 1/60$, $MS_e = 1.53$, $p < .05$), sex of parent ($F = 12.06$, $df = 1/60$, $MS_e = 1.53$, $p < .01$), and Pathology Group \times Sex of Parent ($F = 4.12$, $df = 1/60$, $MS_e = 1.53$, $p < .05$). Comparisons of the four means corresponding to the Pathology Group \times Sex of Parent interaction effect, using t tests, indicates that the mothers of the less pathological group are more incongruent (4.63) than mothers of the more pathological group (3.50) and the fathers of both groups (3.08, each). The other three groups do not differ significantly among themselves.

Analysis of variance of attitude scores based on judgments of attitude in verbalization indicates only one significant effect due to pathology group ($F = 7.77$, $df = 1/60$, $MS_e = .61$, $p < .01$). Attitude communications of the more pathological group (3.05) are less positive than attitude communications of the less pathological group (3.54).

Analysis of variance of attitude scores based on global judgments of attitude in posture indicates no significant effects.

DISCUSSION

The results of the above experiment do not support the hypothesis that there is greater incongruity in the postural verbal communications of parents of more disturbed adolescents compared with the incongruity in the communications of parents of the less disturbed adolescents. Two different measures of incongruity were employed. For the first, global, incongruity measure which was based on direct judgments of the verbalizations and postures of the parents, there is no significant difference in the incongruity of communications from parents of the more and less disturbed groups. For the second, weighted, incongruity measure which is based on anticipated postures in comparison to actual postures for a verbalization, the mothers of the less disturbed group of adolescents are found to show a greater amount of incongruity. In sum, the data fail to support the hypothesis and for one of the measures provide contradictory evidence for the communications of mothers.

The correlations among the various dependent measures provide validation support for the weighting of postural components employed in the present study. It will be recalled that only two significant correlations were obtained; one between judgments of attitude in verbalizations and weighted judgments of attitude in anticipated posture, and another between global and weighted judgments of attitude in actual posture.

The second correlation indicates that the weighted judgments of attitude in posture are comparable to the global judgments of attitude. The first correlation indicates that the quality of anticipated posture in terms of its components is comparable in attitude to the quality of the verbalization upon which the anticipated posture is based.

While the findings involving the incongruity measures are difficult to interpret in terms of double-bind conceptualization, the findings involving measures of degree of positive-negative attitude communication can be interpreted. The parents of the more disturbed adolescents show more negative attitudes towards these adolescents (in their verbalizations, but not in their posture) than

parents of the less disturbed adolescents. An extrapolation of the multicomponent attitude communication model proposed by Mehrabian and Ferris (1967) suggests that the total attitude communicated verbally and posturally is more negative for the parents of the more disturbed adolescents. If the total attitude communicated verbally and posturally can be considered to be representative of the overall quality of attitude communications of these parents toward the adolescents involved, then the findings of the present study suggest a positive correlation between degree of negative attitude communication of parents toward a child and degree of psychopathology of the child.

The relationship between psychopathology of children and the negative attitude communications of their parents can be due to either, or both, of the following. The parents may have more negative attitudes toward these more disturbed children because the children create more problems for them than the less disturbed children do for their parents. Alternately, initially negative attitudes of the parents may have contributed to the psychopathology of the children. In either case, negative attitudes of parents could be considered to at least contribute to the maintainance of the children's maladjusted patterns of functioning. In discussing their findings, Mehrabian and Wiener (1967) suggested that "it could be argued that unusually frequent negative attitude communicating messages do contribute to severe psychopathological functioning. . . . For example, indiscriminate negative reinforcement is not conducive to learning the numerous interpersonal and social skills which are lacking in individuals classed as schizophrenics [p. 114]." Rogers' (1959) conceptualization of psychopathology also suggests a relation-

ship between negative attitude communications of parents and psychopathology of their children. In his theory, greater psychopathology of a child is associated with greater degrees of "conditional positive regard" of parents toward the child. Conditional positive regard refers to the conditional quality of the love or liking of one person toward another, as would, one assumes, be manifest in the frequency of negative attitude communications. In sum, the findings of the present study suggest that the study of the overall quality of positive-negative attitude rather than the incongruity in attitude communication, may be a more helpful avenue for the investigation of the relationship between communication patterns and psychopathology.

REFERENCES

- BATESON, G., JACKSON, D. D., HALEY, J., & WEAKLAND, J. Toward a theory of schizophrenia. *Behavioral Science*, 1956, 1, 251-264.
- HALEY, J. Strategies of psychotherapy. New York: Grune & Stratton, 1963.
- MEHRABIAN, A. Inference of attitudes from the posture, orientation, and distance of a communicator. *Journal of Consulting and Clinical Psychology*, 1968, 32, 296-308.
- MEHRABIAN, A., & FERRIS, S. R. Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 1967, 31, 248-252.
- MEHRABIAN, A., & WIENER, M. Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, 1967, 6, 109-114.
- ROGERS, C. R. A theory of therapy, personality, and interpersonal relationships, as developed in the Client-Centered framework. In S. Koch (Ed.), *Psychology: A study of a science* Vol. 3 New York: McGraw-Hill, 1959.
- WEAKLAND, J. H. The "double-bind" hypothesis of schizophrenia and three-party interaction. In D. D. Jackson (Ed.), *The etiology of schizophrenia*. New York: Basic Books, 1961.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received August 23, 1967)

Oxford University Press



Disordered Behavior:

BASIC CONCEPTS IN CLINICAL PSYCHIATRY

By ERIC PFEIFFER, M.D., *Duke University Medical Center*, "Concise, well written. Deals clearly with some usually hazy areas for students—for example the nomenclature of depression."—Robert L. Leon, M.D., *University of Texas Medical School at San Antonio*

1968

224 pp.

cloth \$5.75

paper \$4.00

Psychological Diagnosis in Clinical Practice:

WITH APPLICATIONS IN MEDICINE, LAW, EDUCATION, NURSING, AND SOCIAL WORK

By BENJAMIN POPE, *The Sheppard and Enoch Pratt Hospital*, and WINFIELD H. SCOTT, *National Institute of Mental Health*. "An excellent introduction to clinical psychological procedures, well illustrated with case material. Suitable for a graduate or undergraduate survey course."—H. Goodglass, *Clark University*

1967

360 pp.

18 illus. (10 in color)

\$8.00

The Experience of Anxiety:

A CASEBOOK

By MICHAEL J. GOLDSTEIN and JAMES O. PALMER, *both of the University of California, Los Angeles*

1963

304 pp.

paper \$2.25

Perspectives in Psychopathology:

READINGS IN ABNORMAL PSYCHOLOGY

Edited by JAMES O. PALMER and MICHAEL J. GOLDSTEIN

1966

448 pp.

44 illus.

paper \$4.25

Foundations of Psychopathology

By JOHN C. NEMIAH, M.D., *Harvard Medical School and Beth Israel Hospital, Boston*

1961

352 pp.

paper \$3.50

OXFORD UNIVERSITY PRESS

200 Madison Avenue, New York, N.Y. 10016

PERSONALITY DEVELOPMENT AND PSYCHOPATHOLOGY

A Dynamic Approach

NORMAN CAMERON

This is an integrated, systematic study of normal and abnormal personality development from childhood to old age. Following six chapters on normal development, the book centers on disorders and reactions. Thorough treatment is given to neuroses and psychoses, with a rich supply of case histories, a description of personality background, and an account of the normal counterpart. The concluding chapter deals with the various forms of therapy in use today, their application, and their history.

791 pages 1963 \$9.95

SOURCEBOOK IN ABNORMAL PSYCHOLOGY

LESLIE Y. RABKIN and JOHN E. CARR

Drawn from both recent research and classic papers in the field, this collection of readings is designed to give the student an understanding of psychopathology. The selections focus on the dynamic elements and family aspects of disturbed behavior, examining such basic questions as: "What is the nature of the interpersonal field of the disturbed person, and out of what kind of family matrix does disturbance arise?" Prefacing each section is a brief introduction to the subject covered and a resume of the articles which follow.

528 pages 1967 Paper \$5.25

A PHILOSOPHY OF SCIENCE FOR PERSONALITY THEORY

JOSEPH F. RYCHLAK

This book provides a framework within which the long-standing problems of psychology can be better understood. Part 1 reviews the background issues which have led to controversy and discusses the nature of theory and method in science. In Part 2 the basic conflict between the "computerized" or mechanistic conception of man and the "humanistic" or dynamic one is shown to have grown out of the historical bifurcation between dialectic and demonstrative reasoning. The author argues that psychologists must learn to consider both points of view in reasoning about man, or run the risk of being incorrect in their conceptualizations or unscientific in their methods of study.

508 pages 1968 \$7.95

HOUGHTON MIFFLIN COMPANY

Boston/Atlanta/Dallas/Geneva, Ill./New York/Palo Alto

*Consider this important
new text . . .*

CASEBOOK ON ETHICAL STANDARDS FOR PSYCHOLOGISTS

The application of the 19 principles comprising the Code of Ethics for all psychologists is presented. Each principle is illustrated by actual, but disguised, cases involving alleged violations of the principle and action taken by the Committee on Scientific and Professional Ethics and Conduct.

Four Appendixes state and amplify the additional principles applicable to specific areas of psychology:

- Ethical Practices in Industrial Psychology*
- Automated Test Scoring and Interpretation Practices*
- Standards for APA Directory Listings of Private Practice*
- Guiding Principles for the Humane Care and Use of Animals*

The arrangement of the casebook into a general section of cases and four additional specific sections increases its versatility as a reference book for individuals and as a text for instructional purposes.

Price: \$1.00

Order From:

**AMERICAN PSYCHOLOGICAL
ASSOCIATION**
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

**FOR THE STUDENT OF
CLINICAL AND ABNORMAL PSYCHOLOGY**
new and forthcoming titles from HRW

THEORY AND RESEARCH IN ABNORMAL PSYCHOLOGY

Edited by David Rosenhan, Swarthmore College,
and Perry London, University of Southern California

The editors consider abnormality in its fullest breadth, including not only what is traditionally called psychopathology, but also such "positive abnormalities" as creativity, intelligence, and constructive social traits.
March 1969 576 pages \$6.95 paper (tent.)

PRINCIPLES OF BEHAVIOR MODIFICATION

By Albert Bandura, Stanford University
July 1969 640 pages \$10.00 (tent.)

METHODS OF MADNESS

The Mental Hospital as a Last Resort

By Benjamin M. Braginsky, Wesleyan University,
Dorothea D. Braginsky, Fairfield University, and
Kenneth Ring, University of Connecticut

The authors make it irrefutably clear that when dealing with even the most mundane activities of mental patients, the dominant psychiatric conception of mental illness is not only inadequate but inappropriate.
April 1969 224 pages \$5.00 paper (tent.)

PSYCHOTHERAPEUTIC AGENTS

New Roles for Nonprofessionals, Parents, and Teachers
Edited with commentaries by Bernard G. Guerney, Jr., Rutgers, The State University

Through articles and editorial commentary, Professor Guerney explores remedial methods and procedures based on the strategy of nonprofessionals as trained and supervised agents of the professional psychotherapist.

March 1969 608 pages \$10.00 (tent.)

HELPING AND HUMAN RELATIONS

A Primer for Lay and Professional Helpers

Volume 1: Selection and Training

By Robert Carkhuff, University of Massachusetts, Amherst

The first of a two-volume series, this volume deals creatively and pragmatically with the effective selection and training of lay and professional personnel for social work, counseling, clinical work, psychiatry, as well as other areas of human relations.

August 1969 288 pages \$7.50 (tent.)

CHANGING PERSPECTIVES IN MENTAL ILLNESS

Edited by Stanley C. Plog, Behavior Science Corporation, California, and Robert B. Edgerton, University of California, Los Angeles

Leading researchers and theoreticians present their beliefs about the causes or contributors to mental illness.

March 1969 768 pages \$15.00 (tent.)



HOLT, RINEHART AND WINTSON, INC.
383 Madison Avenue, New York 10017

EUROPE

A flexible travel program has been developed in connection with the XIXth International Congress of Psychology in London, July 27/August 2, 1969. Several groups will leave New York to Europe, the Mediterranean and the Near East -- utilizing the low group fares of Pan Am, TWA, BOAC and other major airlines. Round trip jet fares start at \$230 (New York/London/New York.) Write for brochure to: Division 'P', Travel Wholesalers International, 1707 L Street, N.W., Washington, D.C. 20036.

Journal of Abnormal Psychology

Diary No. 266.....
Date..... 12-6-77
File No. Lib.....
Bureau Fdnl. Psy. Resear.....

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
*New York State Psychiatric Institute
and Columbia University*

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt

Journal of Abnormal Psychology

Department of Psychology

N. Y. State Psychiatric Institute & Columbia University

722 West 168th St.

New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR	MYRTLE HENDRICKS	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

PSYCHOLOGICAL DIFFERENTIATION AND PERCEPTUAL DEFENSE:

STUDIES OF THE SEPARATION OF PERCEPTION FROM EMOTION¹

JAMES G. MINARD²

AND

WILLIAM MOONEY

State University of New York
Downstate Medical Center, Brooklyn

Western Psychiatric Institute
University of Pittsburgh

Psychological differentiation (field independence, analytical thought) is assumed to involve separation of emotion from perception. Findings supported this assumption as well as previously criticized clinical investigations and an interpretation of perceptual defense. Perceptual measures of differentiation correlated ($p < .025$) with the degree to which emotion altered the probability of tachistoscopic recognition. Recognitions by poorly differentiated Ss included 22.4% fewer emotional than neutral stimuli ($p < .05$). This type of perceptual defense correlated positively, as predicted, with psychological defense serving to blot out emotional experience. Verbal response bias, general accuracy, and verbal intelligence were controlled or unrelated to findings. Discussion emphasized relevance to findings of certain research on the process of stimulus representation and on situations involving rapid response to fleeting stimuli.

Both the critics and *Es* involved have recognized that contamination of variables is a significant problem in study of *psychological differentiation*; that is, study of behavior correlated with the ability to be independent of a perceptual and social field and to think in an analytical way (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Contamination (bias in the evaluation of one variable caused by information regarding another) has been a particularly difficult problem in testing the important assumption that more differentiated people are characterized by greater separation of their perception from their emotion (the separation assumption).

Problems of contamination have not come from indicators of differentiation, since these have included objective perceptual tasks in which an *S* adjusted himself or a rod to a

vertical position or recognized a figure within an embedding context. Rather, the possibility of contamination has arisen when emotional perceptual separation has been evaluated by clinical procedures, since these procedures also permit the evaluation of differentiation. Though attempts have been made to reduce possible contamination, data such as that provided by Rorschach tests, interviews, or diagnoses do not permit convincing elimination of information on differentiation. So long as clinical procedures are used, critical readers may be convinced differentiation could have contaminated all other evaluations and may use this as a basis for rejecting much research on correlates of differentiation (e.g., Zigler, 1963a, 1963b).

Critical test of the separation assumption requires evaluation of emotional perceptual separation by a procedure which is not known to provide information on differentiation. The investigation reported here used an objective, restricted procedure: measurement of visual perceptual defense. Perceptual defense is defined as a description of the relationship between stimulus emotionality and the ease of recognition of stimuli (Brown, 1961; Minard, 1965). In the present study, perceptual defense was indicated by a difference in the probability of accurate responses attributed to the emotional significance of stimuli. The

¹ This study was initiated and all measures of differentiation obtained in the Psychology Laboratory, Department of Psychiatry. Research reported in this paper was supported in part by a grant (M-628) from the United States Public Health Service, National Institutes of Health. The writing of the paper was aided by the reading and comments of Kathleen M. Young, Robert Loiselle, John Dulaney, Gene Ingledue, and Donald Coleman.

² Requests for reprints should be sent to James Minard who is now at the Western Psychiatric Institute, University of Pittsburgh Medical School, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213.

accurate responses were forced choices made as words were repeatedly presented at a threshold level. The perceptual defense difference involved ePa , the probability of accuracy when the presented word was emotion arousing, and nPa , the corresponding probability when the word was nonemotion arousing, but was otherwise similar: perceptual defense = $nPa - ePa$.

The possibility that this difference was caused by nonperceptual variables was experimentally and statistically controlled by previously developed techniques (Minard, Bailey, & Wertheimer, 1965). These techniques and the selection of emotional stimuli are described in the Method section.

Differentiation was evaluated by three frequently used perceptual (or at least nonverbal) tasks (Witkin, 1965). In each task it was advantageous to S to have ready access to a stable representation of some aspect of his previous experience, an aspect such as the feel of being vertical or the shape of a visual stimulus. For the Body Adjustment Test (BAT) he adjusted his body to the gravitational vertical while in a tilted room. For the Draw a Person (DAP) rating scale, he drew a human figure of each sex and was evaluated on the basis of the sophistication of body concept indicated by his drawings.³ For the Embedded Figures Test (EFT) he recognized previously presented geometric figures when they were included as integral parts of patterns.

One hypothesis clearly follows from the separation assumption and the definition of perceptual defense: The less the differentiation of an S , the greater the effect of emotion on perception, and the greater the absolute $nPa - ePa$ difference. Because the separation assumption does not specify whether emotion facilitates or disrupts perception, the hypothesis refers only to the absolute $nPa - ePa$ difference and not to its direction.

³ Drawings were rated by Hanna Marlens, who developed the scale and participated in study of its reliability and validity (Witkin et al., 1962). She saw only the drawings and was not otherwise informed about this study. Other differentiation measures were administered and scored by Judy Birnbaum and William Fishbein.

Additional hypotheses regarding the direction of this difference can be based on previous clinical investigations (Witkin, 1965) and on the assumption that perceptual and psychological defenses are related (Brown, 1961, pp. 68-72). Well-differentiated Ss are thought to "isolate," separating experience from affect, so they may recognize emotional and neutral stimuli with equal ease, suggesting that emotional perceptual separation can be great enough to eliminate the effect of emotion on the recognition of stimuli. Poorly differentiated Ss are thought to "blot out" emotionally significant experience, so they may show less accuracy in responding to emotional stimuli than in responding to more neutral stimuli. Such a result would suggest that, when emotional perceptual separation is too slight, emotion disrupts stimulus recognition. The Ss who are average in psychological differentiation have not been studied enough to allow prediction of their perceptual defense. According to the separation assumption, they have an intermediate separation of perception from emotion, but it is not known whether this will disrupt or facilitate the process of stimulus recognition.

Description of two studies follows. The first used perceptual defense and the three measures of differentiation to test hypotheses implied by the separation assumption, previous clinical investigations, and the assumption that perceptual and psychological defenses are related. The second more directly tested the assumed relationship of perceptual and psychological defense. The Discussion section considers the numerous applications of the concept of emotional perceptual separation, and theoretical interpretation of its occurrence.

STUDY I

Method

The Ss were 30 Brooklyn College males between 18 and 25 yr. of age. They volunteered for a paid "perceptual experiment" announced in their psychology classes and at the student employment office. Previously obtained BAT norms indicated the sample included both well-differentiated and poorly differentiated Ss .

Order of test administration functioned as a random variable since BAT, EFT, and DAP were administered at times determined, not by any S characteristic, but by the availability of research

assistance and apparatus. Two experimental sessions were required for perceptual defense evaluation. During one session BAT and EFT were given. During the other session duration of the two testing periods was equalized by administering DAP and the Information and Comprehension subtests of the Wechsler Adult Intelligence Scale (WAIS). The *Es* evaluating differentiation had nothing to do with the evaluation of perceptual defense and did not know perceptual defense scores.

Perceptual defense. Methods of perceptual defense measurement, and their rationale, have previously been described (Minard, 1965; Minard et al., 1965). During the first of the two sessions, each *S* was given an individually administered word-association test (WAT) containing 42 potentially emotional words selected with the aid of judges familiar with the college population. The judges' task was to select words which were quite familiar yet potentially emotion arousing (e.g., adjust, afraid, child, dance, date, defect, dependent, dumb, failure, love, school, suck, sin, wish, etc.). For each word, there was a presumably neutral word matched as closely as possible (e.g., sit matched with sin). First letter and length were always identical for paired words and they were matched in frequency according to the Thorndike-Lorge count (1944). Matched words had as many corresponding letters as were allowed by the Thorndike-Lorge list. Words were typed in capitals with an electric typewriter. Each was typed in the center of an index card (for a subsequent word-association test) and on the gray, architect's tracing paper used to provide a low-contrast background when stimuli were tachistoscopically presented. The word-association test was administered individually with the instructions "say the first thing which comes to mind, as soon as it comes to mind, when you see the word on the card." Cards were presented at 5- to 10-sec. intervals and responses conspicuously timed with a stopwatch. Four personally emotional words were selected from the list of potentially emotion arousing words. Each selected word had elicited an associative reaction time within the longest 8% of all times for that *S*, a time longer than the average time of other *Ss* on that word. If a word elicited evidence of emotional arousal such as blocking, stammering, long explanations, or sweating, it was chosen in preference to words similar on the basis of other criteria. The four selected words and the neutral words matched with them were arranged in alphabetical order to form the list used for that *S* in the subsequent perceptual defense experiment. Approximately 1 wk. later, *E* familiarized *S* with the selected words, using a slide projector to present them one at a time. Words remained on the screen until identified and *S* was informed he would be asked to recite the list when it was completed. After *S*'s recitation, *E* recited the list rapidly and asked *S* to repeat it.

The list was then presented as *E* adjusted the tachistoscope to a threshold level; that is, a point which allowed identification of the general shape or fragments of the letters of a word whenever *S* was

especially attentive. As a check on the selected conditions, the words were presented once again to see if *S* could identify more than one and less than four words in a series of eight. Stimuli were repeatedly presented at this threshold level during the following measurement trials. Accuracy during these trials was between 36.7% and 45.7%, using 90% confidence limits determined by Student *t*. To permit corrections, a randomly selected half of the stimulus pairs were omitted and smudged blank slides added without *S*'s knowledge. Each stimulus was presented 10 times and 16 blanks were presented during the total of 56 presentations. Each *S* made a forced choice from stimuli on his list after each presentation. Other, less important procedural details have been described previously (Minard, 1965; Minard et al., 1965).

Accurate responses were used to compute the $nPa - ePa$ difference, the perceptual defense score. Responses to blank slides were used to compute an $nPg - ePg$ difference (the difference in response probabilities when responses are guesses). This was the verbal response bias score. Response bias and other information provided by the experiment allowed correction of perceptual defense for non-perceptual variables and more accurate measurement of emotional biasing of the probability of stimulus recognition.⁴

⁴ Corrections are illustrated with accurate neutral responses. In principle, corrections were the same for all stimulus-response categories.

Correction must consider the probability of chance accuracy, response bias, and the tendency of an *S* to restrict his responses to a reasonable set of alternatives. If there is no response bias and *S* chooses freely from the list of eight, the expected number of chance accuracies when a neutral word is presented is equal to the number of guesses to neutral words (n) times the probability of chance accuracy (Pca):
 number of chance accuracies = (n) (Pca).

However, although *S*'s list contains eight words, the probability of chance accuracy is not necessarily $\frac{1}{8}$. Only four words on *S*'s list are presented and only these four can be accurate responses. By restricting his choices to the presented words, *S* may increase his probability of chance accuracy to $\frac{1}{4}$. When *S* chooses from both the presented and the unrepresented stimuli, his probability of chance accuracy is the joint probability of choice from the presented group (Ppg) and the probability of chance accuracy when such choices are made ($\frac{1}{4}$):

$$Pca = (\frac{1}{4}) Ppg. \quad [1]$$

The probability of choice from the presented group is empirically determined from responses to blank slides.

The tendency to either choose or avoid neutral responses (response bias) will change the number of neutral responses accurate by chance. If neutrals are preferred, the number of chance accuracies will be increased; if they are avoided, it will be reduced. The change will be equal to the number of chance

Though the authors believe perceptual defense is sometimes due to verbal response bias, analysis indicated that in this study verbal response bias and other nonperceptual variables did not make a significant contribution to the relationship between perceptual defense and psychological differentiation.

Differentiation. For EFT, *S* was shown a simple geometric figure, asked to describe it, then to locate it in a more complex design. After the first, practice, trial he was given 12 measurement trials with different simple and complex figures. His score was the total number of seconds required to locate figures. For BAT, *S* was seated in a chair facing the wall at the end of a small room. Wood molding and pictures clearly defined the wall's horizontal and vertical aspects. Both room and chair could be tilted independently. During the six experimental trials, the room was always tilted 35° and the chair 22°. Every possible combination of tilts was presented,

accuracies multiplied by a verbal response bias coefficient (*c*). This number is added, algebraically, to (*n*) (*Pca*).

The number of chance accuracies, corrected for response bias is,

$$(n)(Pca) + (c)(n)(Pca). \quad [2]$$

Here, *c* is the difference between the proportion of neutral responses when blanks are presented and .50—the proportion suggesting no response bias. If the difference is negative, algebraic addition will reduce the estimated number of chance accuracies.

Finally, it is false to assume *S* is only guessing throughout the experiment. One must estimate the number of guesses to neutral words (*n*) because *S*'s description of his responses, his accuracy, and the conditions for presentation assures one that he was sometimes accurately recognizing the presented stimulus and was not always accurate just by chance. The number of guesses is not known, but the number of wholly inaccurate responses (*I*) is. This number has been reduced by the number of responses accurate by chance, which is equal to (*n*)(*Pca*):

$$I = n - (n)(Pca). \quad [3]$$

It is a matter of algebra to solve for *n*,

$$n = I / (1 - Pca). \quad [4]$$

This allows the desired correction of the number of accurate neutral responses to obtain a more correct estimate of the number of recognitions. The corrected number of accurate neutral responses (*NAc*) equals the observed number of accurate neutral responses (*NA*) less the expected number of chance accuracies, corrected for responses bias by Formulas 2 and 4:

$$\begin{aligned} NAc &= NA - ((n)(Pca) - (c)(n)(Pca)), \\ &\quad \text{simplifying} \\ NAc &= NA - (1 - c)(n)(Pca). \end{aligned} \quad [5]$$

Further information on the correction procedure has been provided elsewhere (Minard, 1964).

half with room and chair opposite in direction of tilt; half with tilts the same. On each trial, *S* instructed *E* to adjust his chair to a position which seemed vertical to *S*. Vertical was carefully and repeatedly defined as the gravitational vertical. The BAT score was the total degrees of error for all trials. Further information on EFT and BAT has been provided by Witkin et al. (1962) who used the two tests together with the rod-and-frame test as their "perceptual index" of differentiation.

For the DAP rating scale, *S* was asked to "draw a person," then to "draw a person of the opposite sex." Drawings were used by the DAP rater to arrive at a rating based on a number of specified criteria. These criteria used observable characteristics of figures rather than projective interpretations (e.g., one characteristic contributing to a low score is "objectively interchangeable male and female figures"). Criteria falling into the categories of "identity and sex differentiation," "form level," and "level of detailing" were used to assign *S*s to defined points on a 5-point scale. More complete description of the scale and data regarding its reliability and validity have been reported previously (Witkin et al., 1962, p. 117). Reported interjudge correlations for DAP ranged from .83 to .92. Reported correlation with the perceptual index of differentiation was .50. In the present study, correlation of DAP rank with sum rank on BAT and EFT was .63 ($p < .005$, Pearson *r*). Correlation of BAT rank with EFT rank was .45 ($p < .01$).

In summary, differentiation was measured as in previous research. Evidence for the validity of its measurement is provided both by that research and by the significant intercorrelation of tests in the present sample.

Results

Results were consistent with the assumption that greater differentiation is associated with greater separation of perception from emotion. In evaluating differentiation, *S*s were ranked according to scores on BAT, EFT, and DAP; ranks were then averaged across the three tests for each *S*. In evaluating degree of separation between perception and emotion, *S*s were ranked on the basis of their absolute $nPa - ePa$ difference, using accurate-response probabilities algebraically corrected to reduce the importance of nonperceptual variables. For convenience, this score will be termed degree of bias in recognition.

Since high scores on BAT, EFT, and DAP indicate lack of differentiation, there should have been a positive correlation between their ranking (lack of differentiation) and degree of bias in recognition. The Pearson correlation of the two variables was in the predicted

direction and significant (.41, $p < .025$). A Kendall rank-order correlation was significant at the same level and was unchanged by correcting for tied ranks ($\tau = .27$).

Next, the analysis considered the correlation between degree of bias in recognition and each measure of differentiation (Table 1). The correlation of DAP and degree of bias in recognition was also statistically significant as indicated by rank-order correlation ($\tau = .38$, $p < .001$). Neither correlational nor graphic comparisons of the relationship of degree of bias with the three measures of differentiation suggested differences between tests which were statistically significant or theoretically meaningful.

Could the assumed separation of perception from emotion really be due to nonperceptual variables? An advantage of the perceptual defense task is that it permits objective investigation of this question. If reduction of bias in recognition were due to verbal intelligence, a negative correlation between the WAIS subtests and degree of bias would have been found. Instead, summed WAIS scale score had a positive, statistically insignificant Pearson correlation with degree of bias (.126). If the apparent degree of bias in recognition were really due to verbal bias in the responses used to report recognitions, then bias in recognition should have had a fairly high, positive correlation with degree of verbal response bias ($nPg - ePg$, absolute, based on guesses in response to smudged blank slides). Instead, degree of verbal response bias and degree of bias in recognition had a correlation which was small and negative (-.12). The general level of tachistoscopic accuracy correlated only -.04 with degree of bias in recognition. (Correlation of general accuracy with

lack of differentiation was only -.005.) Unless post hoc hypotheses are to be adopted, it must be concluded that the evidence of separation of perception from emotion was not due to nonperceptual variables or to the general level of accuracy.

The direction of the uncorrected $nPa - ePa$ difference was as expected for the well-differentiated and poorly differentiated third of the sample. Poorly differentiated Ss accurately identified fewer emotion-arousing than neutral words: of their accurate responses, 61.2% were neutral and only 38.8% were emotional. The difference of 22.4% is significant ($t = 2.08$, $p < .05$). Well-differentiated Ss recognized about an equal number of neutral and emotional words (53.2% neutral and 46.8% emotional). The 6.4% difference is not significant ($t = .90$). Average Ss showed a "vigilant" tendency to recognize more emotion-arousing than neutral words: 57.0% of accurate responses were emotional and only 43.0% were neutral, a difference of -14% ($t = 2.44$, $p < .05$, two-tailed). Even though Ss who were average in differentiation recognized the highest percentage of emotional words, they were average in terms of the absolute $nPa - ePa$ difference, as predicted by the separation assumption.

Statistics applied to all three group means supported the t tests applied to individual means: analysis of variance showed the groups to differ significantly in perceptual defense ($F = 4.99$, $p < .05$). Duncan's multiple-range test showed perceptual defense of average Ss differed significantly ($p < .05$) from that of the least differentiated Ss. Differences between group means were not found for either general accuracy ($F = 2.208$) or verbal intelligence ($F = .698$). Verbal response bias did not differ significantly from group to group ($F = 2.711$).

Algebraic corrections of the $nPa - ePa$ difference provide the most critical evidence regarding the contribution of nonperceptual variables to group differences in perceptual defense. These corrections consider more than simple verbal response bias and have previously eliminated a substantial $nPa - ePa$ difference created by verbal conditioning (Minard et al., 1965). As the dashed line

TABLE 1

POSITIVE CORRELATIONS BETWEEN LACK OF
DIFFERENTIATION AND DEGREE OF
EMOTIONAL BIAS IN RECOGNITION

	Test		
	EFT	BAT	DAP
Degree of Bias	.37*	.24	.41*

* $p < .025$.

in Figure 1 indicates, the effect of correcting perceptual defense results was very slight. Group differences in perceptual defense remained significant after correction ($F = 5.919$, $p < .05$).

Discussion

Results supported both the separation assumption and the definition of perceptual defense. As predicted, the less the differentiation the greater the effect of emotion on perception, as indicated by an objectively measured perceptual defense task. Had this not been the case, the authors would have had reason to question not only the conception of psychological differentiation, but also the conception that the absolute $nPa - ePa$ difference indicates the degree to which emotion affects stimulus recognition. Because of the way in which emotional perceptual separation was measured here, contamination by clinical judgment can be rejected as an explanation of the results of this experiment and may be

questioned as an adequate explanation of the results of previous work.

The $nPa - ePa$ difference was 6.4% in well-differentiated Ss, -14% in Ss of average differentiation, and 22.4% in poorly differentiated Ss. These results supported hypotheses based on previous clinical studies and on the assumption that there is a relation between perceptual and psychological defense. If the separation assumption is accepted, perceptual defense results also suggest the hypothesis that a low level of emotional perceptual interaction has little effect on recognition, an intermediate level of emotional perceptual interaction increases the probability of recognition, but too high a level of emotional perceptual interaction disrupts recognition of emotional stimuli. The substantial $nPa - ePa$ difference in poorly differentiated Ss, together with previous clinical reports that such Ss "blot out" emotional experience, suggested the desirability of a more direct test of the assumption that perceptual defense is corre-

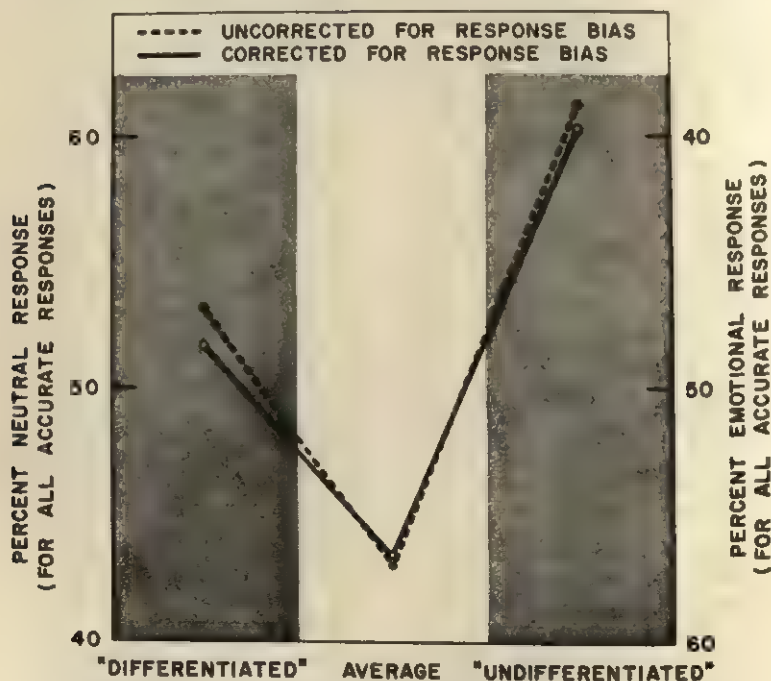


FIG. 1. Correction for nonperceptual variables did not significantly change results demonstrated by the statistical analysis. (The percentages of accurate responses which were neutral or emotional, and their relationship to Ss' differentiation remains essentially as before.)

lated with increased reliance on psychological defenses such as denial and repression.

STUDY II

In this test of the assumed relation between perceptual and psychological defense, an appropriately experienced clinician, uninformed regarding perceptual defense, evaluated college males from whom another *E* obtained perceptual defense scores. The clinician's procedure allowed consideration of the potentially contaminating influence of *Ss*' ability to cooperate. The perceptual defense procedure allowed explicit consideration of the possible contribution of verbal response bias and general accuracy. It was predicted that perceptual defense would be correlated with greater reliance on defenses such as simple repression.

Method

Using the same procedures as in Study I, 33 University of Pittsburgh males were obtained and their perceptual defense determined. Each was seen twice by a psychoanalyst whose previous psychiatric experience had included many evaluations of overtly normal college-age men. His 30-45-min. interviews were tape-recorded. All *Ss* were overtly cooperative, but differed in their ability to respond to questions such as "describe yourself." Immediate impressions and tape recordings were used by the interviewer to rank *Ss* on lack of cooperation during the interview as well as reliance on denial, avoidance, and simple repression.

Results and Discussion

As Table 2 shows, perceptual and psychological ("ego") defense proved correlated as predicted. Psychological defense correlated .45 with lack of cooperation ($p < .005$). However, lack of cooperation did not correlate with perceptual defense. The statistically insignificant correlation between lack of cooperation and response bias may be large enough to warrant further investigation using a perceptual defense procedure in which verbal response bias is encouraged. Prediction of correlation between psychological defense and the *nPa* - *ePa* difference proved to require consideration of the direction of that difference. This finding emphasizes the fact that, although degree of emotional bias contributes to perceptual defense, it can not be equated with it.

TABLE 2

CORRELATION BETWEEN PSYCHOLOGICAL
AND PERCEPTUAL DEFENSE

	Lack of cooperation	Psychological defense
Perceptual defense		
Uncorrected	.09	.41*
Corrected	.08	.43*
Response bias	.19	.10
General accuracy	-.01	-.06
<i>Pn-Pe</i> , absolute	.08	.03

* $p < .01$.

GENERAL DISCUSSION

No matter how convincing the demonstration that emotional perceptual separation is not dependent on contaminated clinical evaluations, no matter how relevant separation may be to psychological defense, explanation of emotional perceptual separation would have to be sought in assumed artifacts if rational scientific explanations were not available, or at least potentially available.

Brown (1961) has described at least six theoretical explanations of perceptual defense, each applicable to the interpretation of emotional perceptual separation. For example, two interpretations are based on assumed mechanisms of psychological defense (pp. 68-72). Though applicable, none of these interpretations is especially suited to data of the present experiment. Available interpretations fail to distinguish clearly between emotionally biased recognition and purely verbal response bias. They give no explicit attention to the concept of psychological differentiation and little attention to current concepts of the process by which stimuli are represented.

Progressive changes in the nonverbal representation of stimuli, accomplished during psychological differentiation, have been implied by theorists such as Hebb (1959, 1963) and Piaget (e.g., Flavell, 1963, p. 56). The authors' research indicates that greater differentiation is accompanied by an increased tendency to form and maintain a stable representation of a naturally fixated stimulus. In this work the authors observed that, for undifferentiated *Ss*, natural fixation can sometimes be enough to make a large, clearly visible stimulus seem to disappear or fragment

for as much as 80% of a 1- or 2-min. fixation period. The typical correlation between a differentiation measure and the duration of stimulus disappearance is about the same as that between differentiation and the degree of emotional bias in recognition (e.g., Minard, 1967; Minard & Batchner, 1967). It is assumed that lack of differentiation is accompanied by a lack of access to stable, well-maintained stimulus representations.

The assumed relationship between differentiation and access to stable stimulus representations may help explain why more differentiated Ss are better able to indicate the gravitational vertical, draw detailed human figures, and identify embedded figures under circumstances which cause less differentiated Ss to perform poorly. Stability of stimulus representations should be explored further by studies involving fixation of a wide variety of stimuli, fixation imposed by special apparatus, and by exploration of working hypotheses suggested by the present results.

The authors hypothesize that lack of stability of stimulus representations may be one factor which increases the degree to which emotional arousal can alter the representation of emotional stimuli. Emotional arousal at any time during the formation, storage, or retrieval of stimulus representations may alter them enough to affect their contribution to the task of recognition. In poorly differentiated Ss, marked alteration of representations of emotional stimuli may interfere with recognition. In Ss of average differentiation, more moderate alteration of representations may actually facilitate recognition (e.g., by reducing the importance of trivial differences between presented and previously experienced stimuli). In well-differentiated Ss, the stability of stimulus representations may be so great that emotional arousal has no significant effect on them. Further refinement of this formulation calls for studies in which emotional arousal is experimentally manipulated while the stability of stimulus representation is continuously measured. To summarize, emotional perceptual separation is open to rational interpretation based on existing theory, but may prove better explained after theoretically oriented research on the rela-

tionship between differentiation and the characteristics of stimulus representations has been carried out.

Emotional perceptual separation is important in interpreting relationships between differentiation and psychopathology. In these interpretations, the separation assumption is used to account for the strength and type of psychological defense. Witkin (1965) has reviewed applications of the concept of differentiation to over 15 types of psychopathology and to the course and outcome of therapy. It would be truly unfortunate if such a range of applications involved an assumption based on contaminated clinical evaluations. In the present research, because of the techniques used, the restriction of data obtained, and restrictions on communication between researchers measuring different variables, contamination of evaluation can be rejected as an interpretation of this study. It may also be questioned as an interpretation of previous work. Though not regarded as successful by all critics, Witkin's group made serious efforts to prevent contamination. They may well have succeeded.

Further research is needed with perceptual defense in samples involving a wider range of psychopathology. However, the present studies of fairly healthy samples probably provided a rather severe test of the hypothesis that perceptual and psychological defense can be related. As with measurement of differentiation, measurement of perceptual defense may contribute to conceptualization of psychopathology. For example, the view that blotting out of emotional experience involves a perceptual process is supported by correlation between perceptual defense and those psychological defenses which serve to eliminate emotional experience. Instead of being attributed to failure of recall, thought, or report, some "blotting out" might best be called, "failure to recognize." Refinement of this conceptualization and better understanding of its implications may result from a more detailed study of the contribution of stimulus recognition to the behavior of a poorly differentiated S.

As described by researchers such as Condon and Ogston (1966) communication in an

interview involves rapid synchronous presentation of more than one type of communication. There is such rapid and frequent response to postural and gestural communication that normal speakers and listeners appear to show "precisely shared patterns of change." These are not evident to casual observation, but become quite evident when behavior is observed with the aid of time-motion analysis of 48-frame-per-second film.⁵ In a poorly differentiated S, perceptual defense against emotional stimuli might interfere with accurate recognition of critical aspects of active communication, partly accounting for the superficial, dependent, anxious quality of his interactions with others. In an S of average differentiation, perceptual vigilance may enhance recognition and response to emotionally significant aspects of communication. This may help him to be the more responsive therapy patient and the more actively interacting therapist described by Witkin (1965) and by Pollack and Kiev (1963). In a well-differentiated S, his capacity for independence, as well as the organization and analysis shown by his thought, may be due in part to a lack of emotional bias in his recognition of stimuli critical to the analysis of experience, including the experience provided by social communication.⁶

In conclusion, the separation assumption is open to rational interpretation, to objective measurement, to use in conceptualizing relations between differentiation and behavior, and to support and application without dependence on potentially contaminated clinical evaluation.

⁵ The description of communication is based on reports by Condon and Ogston (1966) and those they cite. It was checked by a discussion with Condon and Ogston and by looking at their high-speed films, in which interviewer and S often look like two puppets manipulated by the same person.

⁶ Other objective support for relationships between differentiation and defenses has been provided by Witkin et al. (1962) and is not reviewed here. The supplementary relationship of the present study to previous research has also been discussed by Witkin (1965).

REFERENCES

- BROWN, W. P. Conceptions of perceptual defense. *British Journal of Psychology Monograph Supplements*, 1961, No. 35.
- CONDON, W. S., & OGSTON, W. D. Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 1966, 143, 338-347.
- FLAVELL, J. H. *The developmental psychology of Jean Piaget*. New York: Van Nostrand, 1963.
- HEBB, D. O. A neuropsychological theory. In S. Koch. (Ed.), *Psychology: A study of a science*, Vol. 1. *Sensory, perceptual, and physiological foundations*. New York: McGraw-Hill, 1959.
- HEBB, D. O. The semi-autonomous process: Its nature and nurture. *American Psychologist*, 1963, 18, 16-27.
- MINARD, J. G. *The measurement and conditioning of "perceptual defense" and response suppression*. (Doctoral dissertation, University of Colorado) Ann Arbor, Mich.: University Microfilms, 1964, No. 64-9979.
- MINARD, J. G. The response bias interpretation of "perceptual defense": A selective review and evaluation of recent research. *Psychological Review*, 1965, 72, 74-87.
- MINARD, J. G. Frame disappearance and rod-and-frame test performance. *Perceptual and Motor Skills*, 1967, 24, 66.
- MINARD, J. G., BAILEY, D. H., & WERTHEIMER, M. The measurement and conditioning of "perceptual defense," response bias, and emotionally biased recognition. *Journal of Personality and Social Psychology*, 1965, 2, 661-668.
- MINARD, J. G., & BATCHELOR, J. Disappearance of naturally fixated luminous stimuli and stability of phase sequences: A selective review and related experiment. *Perceptual and Motor Skills*, 1967, 24, 747-752.
- POLLACK, I. W., & KIEV, A. Spatial orientation and psychotherapy: An experimental study of perception. *Journal of Nervous and Mental Diseases*, 1963, 137, 93-97.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University Press, Bureau of Publications, 1944.
- WITKIN, H. A. Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology*, 1965, 70, 317-336.
- WITKIN, H. A., DYK, R. B., FATERSON, H., GOODENOUGH, D. R., KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- ZIGLER, E. A measure in search of a theory? *Contemporary Psychology*, 1963, 8, 133-135. (a)
- ZIGLER, E. Zigler stands firm. *Contemporary Psychology*, 1963, 8, 459. (b)

(Received November 16, 1967)

PERSONALITY CHARACTERISTICS OF MALE TRANSVESTITES: III

P. M. BENTLER¹

University of California, Los Angeles

AND

CHARLES PRINCE

*Foundation for Personality Expression,
Los Angeles, California*

A national sample of 181 transvestites replied to the Personality Research Form, an inventory which assesses 20 areas of normal personality functioning. A control sample of 62 Ss was also tested. In comparison with the controls and the test norms, the transvestites presented themselves as more controlled in impulse expression, more inhibited in interpersonal relations, less involved with other individuals, and more independent.

Three major types of male individuals exist who, with some regularity, dress in clothes society considers appropriate to members of the opposite sex (Benjamin, 1966; Stoller, 1968; Walinder, 1967); homosexuals, transsexuals, and transvestites. The transvestite is usually a heterosexual male who enjoys wearing feminine attire, and for whom cross-dressing is usually a habit of long standing. This report is the third of a series of papers dealing with the psychological characteristics of the male transvestite.

Hardly anything is known scientifically about transvestites, as most reports dealing with these persons suffer from several limitations. In the first place, the word "transvestite" is frequently used to refer to members of any of the three groups described above. Second, practically all reports dealing with transvestites are single case studies. Large-scale studies tend to be summaries of cases previously published in the literature, rather than reports of original research of the investigator, based upon his own first-hand clinical or experimental experience with this population (Housden, 1965; Pauley, 1965). Finally, virtually all reported studies have dealt with psychiatrically disabled individuals. The present research, however, investigated the characteristics of transvestites sampled from the community at large. The reader may be referred to the work of Lukianowicz (1959) for a comprehensive

general summary of knowledge regarding transvestism.

The authors recently published two studies dealing with the average personality profile which transvestites generated for a projective test (Bentler, Shearman, & Prince, 1969) and an objective test (Bentler & Prince, 1969). The first study reported the responses of 25 transvestites to the Holtzman Inkblot Test, while the second study was a report of the responses of 176 transvestites to a standard MMPI type psychopathological inventory. Both studies represented investigations of transvestites selected from the community without regard to their psychiatric status, and the results of these studies were consistent in noting that transvestites showed no gross psychiatric symptomatology. Although a possible indication of thought disturbance was found in the projective study, these trends were not verified on the self-report inventory. The main result which was consistent in the two studies indicated that the transvestites were, on the average, slightly more controlled in their impulses than a control or normative group.

The previous studies have dealt quite specifically with the problem of psychiatric symptomatology. A separate but equally important question to be answered is whether transvestites differ on the average from control Ss on nonpsychiatric personality characteristics. The purpose of this investigation is to compare transvestites with control Ss on a number of personality scales which have no particular bearing on abnormal symptomatology, but may well provide insight into the nature of transvestic behavior patterns.

¹Requests for reprints should be sent to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California.

METHOD

Two national samples of transvestites (TV) unselected with respect to psychiatric status were obtained from the files of a magazine published for transvestites by the second author. The first group, a 1966 sample, consisted of 200 subscribers to the magazine selected from the first portion of the alphabetical subscription list. The second group, the 1967 sample, consisted of 125 subscribers selected from the second portion of that list. The Ss were sent a personality inventory by mail and were asked to cooperate anonymously in gathering scientific information on transvestism.

The Personality Research Form (PRF, Jackson, 1967), Form BB, was sent to all individuals.² This test is an inventory of personality scales designed to measure areas of normal functioning such as those described by Murray (1938). Twenty-two traits are assessed by the PRF, as indicated in Table 1. Each trait scale is composed of 20 true-false items, so that the whole inventory contains 440 items.

The original plan of this study called for comparing the responses of TV Ss directly with normative data available for the inventory, but it became apparent that the average age of TV respondents (43 yr.) was greater by over 20 yr. than the age of the normative sample. Thus, a control group of 62 participants of a University of California extension course was obtained. These persons were "middle-manager" individuals sent by their companies to participate in sensitivity training with the goal of enhancing their business potential. This control sample, although not in any sense an ideal random sample of males in the community, showed approximately the same age and education distribution as the transvestite sample.

RESULTS

The 115 TV Ss in the 1966 sample, 66 TV Ss from the 1967 sample, and all 62 Ss from the control sample returned usable data. Statistics were computed separately for the two TV samples. The means and standard deviations on all scales were remarkably similar for the two TV samples, so the data were therefore combined for purposes of analysis. The left-hand column of Table 1 lists the 22 PRF scales. The first 20 scales describe various aspects of personality functioning, while the last two scales—the Infrequency and Desirability scales—are control scales designed to assess bias in responding. The columns labeled 1, 2, and 3 report the average score on the various scales obtained for three groups

of Ss. The first column presents the means for the 181 TV Ss; the second column reports the scale means obtained with normative data based on 1,029 Ss (norm); the third column represents the mean scores obtained by the control group (control). The differences between TV and norm sample means on the various scales appear in Column 4, with a negative sign indicating that the norm mean was higher than the TV mean, while the absence of a sign indicates the reverse. Column 5 presents the differences between the TV and control sample means. Finally, the standard deviations for the TV and control Ss are presented in Columns 6 and 7.

A discriminant function analysis comparing the control and TV samples showed that the two sets of means differed significantly. The *F* ratio ($df = 22/220$) was 5.45, significant at the .001 level, indicating the average scores obtained by TV Ss differ from those obtained by the controls. Similarly significant scale-by-scale differences were found in comparing TV Ss with the normative sample.

DISCUSSION

Both TV and non-TV Ss showed acceptable degrees of purposeful responding, as indicated by the control scales. The Infrequency scale, which assesses implausible or pseudorandom responding shows that TV Ss respond more purposefully than either set of control Ss. The Desirability scale, which assesses the extent to which Ss present a favorable picture of themselves, indicates that TV Ss do not differ from controls. Thus, the data are sufficiently meaningful to warrant interpretation.

Various groups of PRF scales consistently differentiate TV from non-TV Ss. One set of scales show that TV Ss are clearly more controlled in their impulses than are non-TV Ss. Thus, they obtained higher mean scores on the Harmavoidance and Order scales, and lower scores on the Change and Impulsivity scales. Harmavoidance refers to a tendency to avoid risks of bodily harm and that of not enjoying exciting activities which might involve danger. High scores on the Order scale indicate a concern with keeping personal effects and surroundings neat and organized. On the other side, the relatively low Change

² The authors thank D. N. Jackson for making available his copyright test for distribution in the manner indicated prior to its publication.

TABLE 1
STATISTICAL DATA FOR PERSONALITY RESEARCH FORM SCALES

Scale	(1) TV <i>M</i>	(2) Norm <i>M</i>	(3) Control <i>M</i>	(4) TV-norm <i>M</i> Difference	(5) TV-control <i>M</i> Difference	(6) TV <i>SD</i>	(7) Control <i>SD</i>
1 Abasement	6.18	6.22	5.37	-.04	.81	2.83	2.13
2 Achievement	12.51	12.58	14.65	-.07	-2.13	3.52	3.01
3 Affiliation	12.39	14.98	12.98	-2.59	-.59	4.46	3.93
4 Aggression	4.83	7.93	7.10	-3.10	-2.27	3.39	2.99
5 Autonomy	10.71	8.62	9.94	2.09	.78	3.26	3.07
6 Change	10.81	11.74	12.39	-.93	-1.57	3.08	2.45
7 Cognitive Structure	10.69	10.90	10.66	-.21	.03	3.41	3.09
8 Defence	8.26	8.76	7.69	-.50	.57	2.58	2.54
9 Dominance	8.81	11.07	13.15	-2.26	-4.33	4.49	3.85
10 Endurance	13.12	10.67	13.47	2.45	-.35	3.32	3.32
11 Exhibition	8.01	10.83	10.31	-2.82	-2.30	4.34	4.31
12 Harmavoidance	12.35	7.46	9.97	4.89	2.39	3.98	3.72
13 Impulsivity	7.34	9.78	8.34	-2.44	-1.00	3.87	3.79
14 Nurturance	11.60	12.68	12.69	-1.08	-1.10	3.57	2.98
15 Order	13.20	10.81	12.44	2.39	.77	4.13	3.87
16 Play	9.23	12.13	9.03	-2.90	.19	3.66	3.02
17 Sentience	14.44	15.24	14.90	-.80	-.46	3.03	3.28
18 Social Recognition	8.89	11.90	10.89	-3.01	-2.00	3.48	3.21
19 Succorance	6.97	7.88	7.60	-.91	-.62	3.81	2.95
20 Understanding	12.16	12.46	12.52	-.30	-.36	4.03	3.26
21 Infrequency	.35	.69	.66	-.34	-.31	.63	1.10
22 Desirability	15.82	15.45	16.26	.37	-.44	2.88	2.62

Note.—TV = transvestite.

scores indicate a relative dislike for new and different experiences, and a preference for routine. Low Impulsivity scores are indicative of tendencies to act primarily with deliberation rather than on the spur of the moment, and the tendency to withhold feelings and wishes rather than giving these free expression. These results are consistent with those reported previously by the authors (Bentler & Prince, 1969; Bentler, Shearman, & Prince, 1969) and others (e.g., Randell, 1959).

A second set of differences occurs with regard to the degree and nature of interpersonal orientation. Thus, TV Ss tend, on the average, to be relatively withdrawn from social involvement. Thus, they have lower scores on the Affiliation, Exhibition, Nurturance, and Social Recognition scales. The relatively low Affiliation scores indicate that, in comparison to controls, TV Ss do not as readily enjoy being with friends and people in general, that they do not tend to make efforts to win friendships and maintain associations as readily. The relatively lower scores on the Exhibition scale indicates tendencies to introversion, so that TV Ss would prefer not to be the center of attention and do not report

themselves as enjoying behavior such as witticism or dramatics which would require an audience. The Nurturance scale indicates that TV Ss tend to be less interested in readily performing favors for others, giving assistance, sympathy, and comfort to others who may need it. Finally, the lowered Social Recognition scores indicate that TV Ss report themselves to be less approval seeking in the sense of being concerned about what other people might think of them; they are not quite as concerned as non-TV Ss in desiring to be held in high esteem by acquaintances, nor in working for the approval and recognition of others. These data are consistent with those of Morgenstern, Pearce, and Rees (1965), who noted that their 19 TV Ss were, on the average, one standard deviation below the normative mean on the Maudsley Personality Inventory Extraversion scale.

Similarly, in the area of interpersonal relations, but more particularly dealing with the degree of ascendancy demonstrated by Ss, TV Ss report themselves as relatively lower in Dominance and Aggression. Thus, they attempt to influence and direct other people,

express opinions, and enjoy a leadership role relatively less than control Ss. Similarly, the relatively lowered Aggression scores indicate that TV are not as easily annoyed as non-TV Ss, do not particularly enjoy combat and argument, and are not willing to hurt other people in order to get even with those who they perceive as having harmed them.

Finally, in regard to the willingness to follow directions from other people, TV Ss tend to be relatively more self-reliant and rebellious than others. They report themselves as relatively higher in Autonomy and lower in Succorance. This means, according to Jackson, that TV Ss are relatively more interested in being free, not tied to people or obligations, and that they enjoy breaking away from restraints or restrictions. The relatively lowered Succorance scores indicate that, in comparison to non-TV, TV Ss on the average do not as readily seek the sympathy, protection, love, advice, or reassurance of other people. They report themselves as not feeling insecure when without support by others, and tend not readily to confide their difficulties to another receptive person.

The interested reader can consult the *Personality Research Form Manual* (Jackson, 1967) if he wishes a more extensive description of the various scales for further clarification of the remarkable and consistent differences observed between TV and non-TV Ss.

REFERENCES

- BENJAMIN, H. *The transsexual phenomenon*. New York: Julian Press, 1966.
- BENTLER, P. M., & PRINCE, C. Personality characteristics of male transvestites: II. *Archives of General Psychiatry*, 1969, in press.
- BENTLER, P. M., SHEARMAN, R. W., & PRINCE, C. Personality characteristics of male transvestites. *Archives of General Psychiatry*, 1969, in press.
- HOUSDEN, J. An examination of the biologic etiology of transvestism. *International Journal of Social Psychiatry*, 1965, 11, 301-305.
- JACKSON, D. N. *Personality Research Form Manual*. Goshen, New York: Research Psychologists Press, 1967.
- LUKIANOWICZ, N. Survey of various aspects of transvestism in the light of our present knowledge. *Journal of Nervous and Mental Disease*, 1959, 128, 36-64.
- MORGENSTERN, F. S., PEARCE, J. F., & REES, W. L. Predicting the outcome of behaviour therapy by psychological tests. *Behaviour Research and Therapy*, 1965, 2, 191-200.
- MURRAY, H. A. *Explorations in personality*. Cambridge: Harvard University Press, 1938.
- PAULEY, I. B. Male psychosexual inversion: Transsexualism. *Archives of General Psychiatry*, 1965, 13, 172-181.
- RANDELL, J. B. Transvestism and trans-sexualism. *British Medical Journal*, 1959, 5164, 1448-1452.
- STOLLER, R. *Sex and gender*. New York: Science House, 1968.
- WALINDER, J. *Transsexualism*. Copenhagen, Denmark: Munksgaard, 1967.

(Early publication received November 13, 1968)

REJOINDER TO HIRT AND KURTZ' "A REEXAMINATION OF THE RELATIONSHIP BETWEEN BODY BOUNDARY AND SITE OF DISEASE"

SEYMOUR FISHER¹

AND

SIDNEY E. CLEVELAND

*State University of New York
Upstate Medical Center, Syracuse*

Veterans Administration Hospital, Houston, Texas

This paper is a reply to criticisms made by Hirt and Kurtz of the Fisher-Cleveland exterior-interior body scheme model. It was shown that these criticisms are contradicted by the pertinent literature. Questions were also raised concerning the rationale of the Hirt and Kurtz study, the characteristics of the experimental populations, and the appropriateness of some of the statistical procedures.

Hirt and Kurtz (1969) make a number of theoretical criticisms of the Fisher-Cleveland (1969) work and also purport to challenge the general validity of the exterior-interior model on the basis of a single empirical study. The present authors suggest that many of their statements represent a puzzling lack of acquaintance with the pertinent literature. Further, possible methodological difficulties with their one empirical study are pointed out.

By way of general criticism, they comment that the reliability and validity of the Fisher-Cleveland boundary scoring system have never been established; that the independence of the Barrier score is in doubt because it is positively correlated with the Rorschach $F+$ index; that the relations of the boundary scores to measures of psychopathology have been inadequately explored; and further that their correlations with chronicity and degree of disability associated with illness are not known. However, the existing literature plainly contradicts these assertions.

Thus, with regard to the reliability of the scoring system, there are now many studies which have demonstrated adequate scorer and test-retest reliability for the boundary scores. This information is available to any interested person in the reports of McConnell and Daston (1961), Williams and Krasnoff (1964), Fisher (1963), Fisher and Cleveland (1968), Ramer (1963), Landau (1960), and others.

¹ Requests for reprints should be sent to Seymour Fisher, Department of Psychiatry, Upstate Medical Center, 766 Irving Avenue, Syracuse 10, New York.

The statement by Hirt and Kurtz that the validity of the boundary scores has not been shown is easily challengeable. Apparently, they are not aware that there have been about 125 studies dealing with this question. As Fisher and Cleveland (1968) and also Fisher (1963) have shown, the overwhelming majority have supported the construct validity of the boundary scores. Apropos of this point, note that in a recent review of the literature concerned with the psychosomatic implications of the exterior-interior model, Buss (1966) remarked upon its high predictive success, "This is an excellent record for any theory in the area of psychosomatic disorders [p. 421]."

The Hirt and Kurtz belief that Barrier is "clearly related to" $F+\%$ is also erroneous. If one turns to actual results, one observes that Compton (1964) was not able to find such a relationship. Further, one of the present writers has observed largely chance correlations between Barrier and $F+\%$ in numerous samples of Ss (Fisher, in press).

Hirt and Kurtz state that the relationships of the boundary scores to degree of disability and chronicity of disability are not known; but this represents again a lack of knowledge of the literature. Multiple studies have shown that the correlations in question are of a chance order (e.g., Allardice & Dole, 1966; Fisher & Cleveland, 1968; Landau, 1960; Sieracki, 1963; Ware, Fisher, & Cleveland, 1957).

Another point made by Hirt and Kurtz is that inadequate attention has been directed

to ascertaining the correlation between the boundary scores and ego strength. In actuality, much energy has been devoted to this problem. It is particularly pertinent that neither Holtzman, Thorpe, Swartz, and Herron (1961) nor Fisher and Cleveland (1968) have been able to find consistent correlations between the boundary scores and paper and pencil tests of ego strength, such as the MMPI. The correlation between the boundary scores and an MMPI index of psychopathology which Hirt and Kurtz observed (but only in a small subsample) stands as an isolated observation. Further, the MMPI index they employed has its own serious validity problems. Goldberg (1965), who devised the index, stated in a study in which he examined its utility, "The validity of even the best of the signs reported in the present study may provide no reassuring balm for MMPI users [p. 24]." But more importantly, if Hirt and Kurtz had examined papers by Fisher (1964a, 1966) dealing with the boundary and psychopathology, they would have discovered that much thought has been given to the problem and that currently it is not feasible to assert that the state of the boundary and degree of psychopathology are consistently linked.

Critical questions can be raised concerning the statistical procedures used by Hirt and Kurtz. The chief defect relates to the fact that the assumption of homogeneity of covariance was substantially violated with reference to the Penetration comparison they reported. The fact that there was a negative correlation ($-.626$) between Penetration and the MMPI index of pathology in the interior group and a positive one ($.315$) in the exterior group (a difference of considerable magnitude) makes use of covariance in this case invalid. Further, one must challenge the appropriateness of an analysis of variance design which grossly combines the two exterior groups despite the fact that their means obviously indicate that they differ (and probably significantly) from each other for both Barrier and Penetration. Also, reported mean values indicate there might even be significant differences for Barrier and Penetration between the dermatitis group and each of the interior groups.

The prime virtue that Hirt and Kurtz ascribe to their study is that it is the first to compare more than one type of exterior and interior symptomatology group simultaneously. This is untrue. The original Fisher and Cleveland (1955) monograph demonstrating boundary differences between exterior and interior symptom groups included patients with rheumatoid arthritis, dermatitis, and muscle difficulties in the exterior sample and patients with stomach ulcers and ulcerative colitis in the interior sample. The numbers of different categories of patients were greater than those presented by Hirt and Kurtz in their study. This fact seems to be unknown to them. It is also striking that nowhere in their paper do they mention the large number of studies which have specifically verified the Fisher-Cleveland symptomatological model. These studies cannot be dismissed simply because they usually involve comparing two symptom groups at a time. Logically, a test of the exterior-interior model may be said to occur whenever an exterior and an interior symptom group are compared; and the fact is that these studies have involved a greater diversity of patients than represented in the Hirt and Kurtz samples.

Hirt and Kurtz have also failed to mention in their paper that a large amount of literature (e.g., Williams & Krasnoff, 1964) has accumulated which indicates that the Fisher-Cleveland exterior-interior model successfully predicts differential physiological reactivity of exterior and interior body sites in both normal Ss and patients with "psychosomatic" symptoms. These studies are of crucial importance because they support the model's basic statements about outside versus inside physiological reactivity in a more controlled and specific fashion than can studies based on groups with gross symptomatological differences.

In short, the general criticisms made of the Fisher-Cleveland work by Hirt and Kurtz are contradicted by the real facts. One must question why they have chosen not to mention the succession of studies that do buttress the validity of the exterior-interior model. They have presented their own work as if it existed in a vacuum.

It is ironic that while they criticize the previous Fisher-Cleveland work for not ade-

quately defining the characteristics of Ss used, they present no information at all about their Ss, except to indicate their gross diagnostic labels. This could be a serious defect because it is known that both sex (Fisher, 1964b) and achievement motivation (e.g., as represented by socioeconomic status, Landau, 1960) affect Barrier level. How do we know that the inability of Hirt and Kurtz to find boundary differences between the symptomatological groups did not reflect the fact that these groups varied significantly in sex or socioeconomic composition? Even more confusing is the definition offered of the exterior, interior, and mixed Ss who were described by Hirt and Kurtz as involved in Experiment II. We are told that these Ss were assigned to the groups by a "physician experienced in psychiatry and internal medicine." What do we know about the reliability of this physician's judgments? Also, they remain relative meaningless unless we are given details of the specific symptoms that he included in each group.

If one were to assume there were no serious errors in the way Hirt and Kurtz assigned Ss to the diagnostic groups, what would their findings mean? Since the exterior symptomatological groups (rheumatoid arthritis and dermatitis) employed have already shown themselves in past studies to conform well to the exterior-interior model and since the two interior groups (heart pathology and asthma) have not previously shown such conformance, it would be logical to conclude that these latter groups do not "fit" the exterior-interior schema. But it should be noted that of all of the interior symptom groups studied these are the only ones that would apparently not conform to expectation.

This could represent a limit on the generality of the exterior-interior model. But one might wonder why coronary pathology and asthma should prove to be such exceptions. With respect to coronary pathology, there are multiple factors (e.g., smoking, diet, exercise) which are suspected of playing a direct contributory role. If smoking or diet do play such roles, it would obviously not be fair to expect the exterior-interior model to apply. It would be just as reasonable to expect the exterior-interior concept to apply in any other situa-

tion where an individual is the target of an environmental pathogenic or traumatic force. If a man swallows a poisonous substance and it damages his stomach he could not meaningfully be said to have developed an "interior" symptom. One must ask whether there is any reasonable limit on the range of body symptoms to which the exterior-interior concept can be applied. If it can be applied to coronary pathology, why not to liver damage or kidney pathology? Perhaps it is a defect of the exterior-interior model that it does not define the limits of its applicability in this sense. One can only say that this is not merely the fault of the model. A large share of the difficulty arises from the fact that so little is known about the etiology of many syndromes like stomach ulcers, colitis, and arthritis. However, it does not seem unreasonable to exclude the application of the exterior-interior model to syndromes in which there is evidence that toxins, infections, and trauma arising from "outside" forces play a direct causative role. "Coronary defects" may fall into this inappropriate category.

An explanation of the negative findings with regard to the asthmatic group is not apparent. Perhaps they illustrate a failing in the exterior-interior model. Perhaps they are a function of the fact that asthma may have a strong hereditary component, as asserted by Leigh, Marley, and Braithwaite (1967) and Schwartz (1952). Perhaps they are related to the fact that many asthmatics are on a regimen of sympathomimetic drugs which are known to produce neuromuscular sensations which may, as shown by Fisher and Renik (1966), produce a reinforcement of boundary sensations and therefore artificially increase the Barrier score. This remains to be seen. Incidentally, if Hirt and Kurtz were interested in reassessing previous results dealing with the exterior-interior model, it seems illogical that they should have chosen to represent the interior category with two symptom syndromes that Fisher and Cleveland have never labeled as appropriate for this purpose.

The attack of Hirt and Kurtz upon the Fisher-Cleveland model involves essentially two different components. There is, first of all, a series of broad denunciatory statements which, as shown, are untenable in the face of

even a moderate knowledge of the pertinent literature. Second, there is a single empirical study that omits important details concerning Ss employed; uses inappropriate statistical procedures; and embraces one symptomatological group (viz., coronary defects) that may be quite inappropriate for testing the exterior-interior model. Finally, even if this empirical study were to turn out to be sound, it represents one of a few minority negative findings among many more positive findings by other investigators.

REFERENCES

- ALLARDICE, B. S., & DOLE, A. A. Body image in Hansen's Disease patients. *Journal of Projective Techniques and Personality Assessment*, 1966, 30, 356-358.
- BUSS, A. H. *Psychopathology*. New York: Wiley, 1966.
- COMPTON, N. Body image boundaries in relation to clothing fabric and design preferences of a group of hospitalized psychotic women. *Journal of Home Economics*, 1964, 56, 40-45.
- FISHER, S. A further appraisal of the body boundary concept. *Journal of Consulting Psychology*, 1963, 27, 62-74.
- FISHER, S. Body image and psychopathology. *Archives of General Psychiatry*, 1964, 10, 519-529. (a)
- FISHER, S. Sex differences in body perception. *Psychological Monographs*, 1964 78(14, Whole No. 591). (b)
- FISHER, S. Body image in neurotic and schizophrenic patients. *Archives of General Psychiatry*, 1966, 15, 90-101.
- FISHER, S. *Body experience in fantasy and behavior*. New York: Appleton-Century-Crofts, in press.
- FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychological Monographs*, 1955, 68(17, Whole No. 402).
- FISHER, S., & CLEVELAND, S. E. *Body image and personality*. (Rev. 2nd ed.) New York: Dover Press, 1968.
- FISHER, S., & RENIK, O. Induction of body image changes. *Journal of Projective Techniques and Personality Assessment*, 1966, 30, 429-434.
- GOLDBERG, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychoses vs. neuroses from the MMPI. *Psychological Monographs*, 1965, 79(9, Whole No. 602).
- HIRT, M., & KURTZ, R. A reexamination of the relationships between body boundary and site of disease. *Journal of Abnormal Psychology*, 1969, 74, 67-70.
- HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality*. Austin: University of Texas Press, 1961.
- LANDAU, M. F. Body image in paraplegia as a variable in adjustment to physical handicap. Unpublished doctoral dissertation, Columbia University, 1960.
- LEIGH, D., MARLEY, E., & BRAITHEWAITE, D. *Bronchial asthma*. London: Pergamon Press, 1967.
- MCCONNELL, O. L., & DASTON, P. G. Body image changes in pregnancy. *Journal of Projective Techniques*, 1961, 25, 451-456.
- RAMER, J. The Rorschach barrier score and social behavior. *Journal of Consulting Psychology*, 1963, 27, 525-531.
- SCHWARTZ, M. *Heredity in bronchial asthma*. Copenhagen: Munksgaard, 1952.
- SIERACKI, E. R. Body image as a variable in the acceptance of disability and vocational interests of the physically disabled. Unpublished doctoral dissertation, State University of New York at Buffalo, 1963.
- WARE, K. E., FISHER, S., & CLEVELAND, S. E. Body-image boundaries and adjustment to poliomyelitis. *Journal of Abnormal and Social Psychology*, 1957, 55, 88-93.
- WILLIAMS, R. L., & KRASNOFF, A. G. Body image and physiological patterns in patients with peptic ulcer and rheumatoid arthritis. *Psychosomatic Medicine*, 1964, 26, 701-709.

(Received December 20, 1968)

REALITY TESTING BY SCHIZOPHRENICS

EDWARD A. JEROME,¹ MARGUERITE L. YOUNG²

AND SUSAN BOSTWICK

*National Institute of Mental Health
Bethesda, Maryland*

Fifteen schizophrenics were compared with 15 controls of similar mental test status and educational background on two classes of elementary problems from an extensive series of artifact testing tasks monitored by an automatic problem programming machine called the Heuristic Evaluation Problem Programmer. In terms of an index that reflects primarily the ability to conduct the required inquiry efficiently, the groups did not differ significantly on statistical tests that employed between-Ss differences as the error variance. Repeated measures of group differences did, however, show a small, but statistically significant, superiority for the control Ss. Intercorrelations between scores on various problem sets indicated that the tests were reliable. During the first three-quarters of the problem series, there was a marked inefficiency of performance in a small group of paranoid schizophrenics as compared with a group of chronic undifferentiated schizophrenics.

The schizophrenic's strange behavior, his indefensible beliefs, and the bizarre non sequiturs by which he attempts to rationalize them have created and supported a firm conviction that there is something wrong with the way schizophrenics think. It has not been easy, however, to decide what this deficit is. The logical oddities that abound on the ward, especially those associated with the fourth syllogistic figure, apparently caused von Domarus (1946) to assert that schizophrenic thinking is guided by a "paralogic." They also led Arieti (1955, 1959), in extending this thesis, to describe them as "paleologic," in order to make explicit reference to the phylogenetic regression that he felt to be characteristic of the schizophrenic's aberrant thought processes. This clinical emphasis on logic and syllogisms constituted a definite theoretical proposal that could not be left unchallenged. Several investigators (Gottesman & Chapman, 1960; Nims, 1959; Williams, 1964) have compared the performances of normals and schizophrenics on multiple-choice tests of the ability to differentiate be-

tween valid and invalid syllogisms, but the anticipated deficit did not appear. Schizophrenics recognized logical fallacies at least as well as their controls did, and, consequently, no evidence was generated to support the notion that their aberrant way of thinking might be a result of a rejection of, or regression from Aristotelean principles.

Does this outcome suffice to discredit the hypothesis that there is something wrong with the logical processes of schizophrenics? The tests in question, though favorites of some logicians, fail to reflect the important fact that syllogisms are useful only when viewed as schemata designed to assist in evaluating deductions that are intermediate products in the daily struggle to reconstruct an adequate image of the world. It may have been precisely this functional use of logic that von Domarus and Arieti meant to impugn in their patients. Concern with this interaction between the mind and reality is prominent in several disciplines that deal with the behavior, errant or typical, of human beings. The clinician stresses its importance by the diagnostic value he places on "reality testing" (Weiner, 1966). Miller, Galanter, and Pribram (1960) emphasize its significance for the theoretical psychologist by the determining influence they assign to the image of reality in goal directed thinking. The philosopher has certainly endowed this problem with more devoted contemplation

¹ Requests for reprints should be sent to Edward Jerome, Section on Higher Thought Processes, Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland 20014.

² The authors wish to express their gratitude to Paul Wender and to William Pollin and his associates for their psychiatric evaluation of the subjects, and to John J. Bartko for his advice on statistical considerations.

than any other professional intellectual. John Dewey (1938) essentially equates the ability to think effectively with the ability to conduct competent inquiry into existing states of affairs. This includes the ability to extract valid information from transactions with well-behaved segments of reality as well as the ability to organize and direct an inquiry toward a selected goal. It is the function of logic to supply conceptual paraphernalia propaedeutic to competent inquiry, but here, presumably, its responsibility ends. How the man on the street uses the products of their effort is not considered a proper topic of concern for logicians. Nonetheless, it is in this task of application that schizophrenics and the authors fear, many others may frequently fail, in spite of their ability to recognize invalid syllogisms in multiple-choice tests. Some empirical discipline, perhaps psychology, must undertake to study the human use of the principles of inquiry as they are applied in attempts to control and comprehend the puzzling events that characterize the real world.

Indeed, initial efforts have already been made in this direction (Blatt, 1961; John, 1957; Rimoldi & Devane, 1961), but the experimental tasks that have been developed are rather complicated for some large and important segments of the population (Jerome, 1962; Young, 1966). In the interest of extending the perspective of research on reality testing, it is necessary to develop a variety of problem situations, especially at the lower end of the difficulty scale if one wishes to secure information about the heuristic behavior of people less gifted than college graduates. The Heuristic Evaluation Problem Programmer (HEPP), the test device employed in the present study, was designed with just this purpose in view; it is a well-behaved segment of reality, an artifact to be sure, that *S* investigates under the constraints of a particular procedure. A priori consideration suggests that the problems it monitors have three important properties: (a) they constitute a family in the sense of requiring for their solution the application of a variety of closely related logical forms; (b) they are arranged in an approximate order of difficulty based on objective mea-

sures of complexity (see following section); and (c) they represent a range of difficulty that is quite large and includes at least some classes accessible to the heuristically ungifted.

It was the purpose of the present study to determine whether the procedure described in the following section is capable of eliciting from schizophrenics performances that are about as good as those of control *Ss* of comparable mental test status and educational background on the elementary problems of the HEPP series. It is especially important to succeed in this connection, because the more difficult problem classes are developed from this level by gradual increments of conceptual complexity without introducing procedural changes or new materials. Success on the first two classes of problem, therefore, provides assurance that *S* understands the basic requirements of the situation and is capable of operating adequately in the test environment.

METHOD

Subjects

Fifteen experimental *Ss* with the following sub-classification of schizophrenia were included in the study: chronic undifferentiated, nine; chronic paranoid, four; schizo-affective, one; schizoid personality, one. The duration of illness from the time of first hospital admission ranged from 1.5 to 11 yr., with a mean of 4.6. According to parental reports, corroborated by hospital records, at least seven of these patients had serious personality and social difficulties necessitating placement in special schools or outpatient psychotherapy from the age of 7 on. Fifteen normal controls were selected from a group of National Institutes of Health volunteers who showed no pathology under extensive psychiatric examination. In making this selection, it was necessary to sacrifice age and sex matching in order to achieve approximate equality of mental test scores and duration of formal education. Indexes referring to these extraexperimental variables are shown in Table 1 for the two groups.

Apparatus

The technical details of the apparatus have been described at length elsewhere (Jerome, Young, & Umberger, 1964), but the problem situation can be explained in terms of *S's* operation panel shown in Figure 1. Arranged in a 4 × 4 matrix are 16 letter-labeled indicator switches which constitute *S's* inputs to the system. When pressed, they are illuminated and close signal circuits used by the control unit. On any given trial *S* may operate any number of these buttons, including all of them or none of

TABLE 1

WAIS SCORES, EDUCATION, AND AGE OF SUBJECTS

Item	Schizophrenics		Controls	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
WAIS				
Verbal	108.2	14.3	108.5	7.5
Performance	100.6	17.6	102.9	8.7
Full Scale	105.3	15.4	106.4	7.2
Education	12.1	3.0	11.8	2.8
Age	24.5	6.3	35.3	15.3

them. The button labeled "Tell-Me" (TM), just below the input matrix, is the only system output accessible to *S*. It is an information signal, which, when pressed, is illuminated by a red or a green light depending upon the relation between the states (on-off) of the inputs and the problem constraints effective at that time. A trial consists of a selection of inputs terminated by a pressing of the TM button. The panel, with *S*'s selection illuminated and the information color displayed, remains lit for 3 sec. after TM is pressed. During the display period,

codes representing the selection and the color of TM are recorded on a punched tape for subsequent input into a computer program. The "cancel" button is available at all times, and the numeric readout (top center) provides a running count of the trial number. The three dark buttons (to the left and right of TM, and upper right-hand corner) were not used in this experiment. The *S* had ample opportunity to become familiar with these operating characteristics during the detailed instructions for the first problem class.

Instructions

All instructions were given on an eight-light field, the two top rows being covered by an opaque screen. Immediately after the instructions on a given class were completed, *S* solved five problems on this same field. The *E* made a maximum effort to convey to *S* all information relevant to each problem class as it was introduced. There were certain operating options and constraints common to both classes. In any problem there was only one input that had anything to do with the color of TM. Before this input had been identified in a given problem, it was referred to as "HEPP" merely as an expository stratagem to avoid the awkward repetition of "the one and only light that has anything to do with the color of Tell-Me." The *S* was always free to turn on any number of lights. The procedure was completely *S* paced; *S* was encouraged to take his time to plan. It was always *S*'s task to find HEPP with as few TM presses as possible. Within any problem class, a change of problem consisted in changing the particular input that influenced the color of TM. The *S* was elaborately informed about every change of problem. The instructions employed to convey this information were highly redundant and consisted of a sequence of short, simple statements, each demonstrated, when possible, by panel operations executed by *S* at the direction of *E*. The *S* was encouraged to ask questions, and *E* repeated instructions and demonstrations until he was convinced that the problem was understood as clearly as possible under the circumstances that prevailed. If necessary, explanations were continued during *S*'s first solution efforts.

With this background the two problem classes used are easily defined: In HEPP I problems there was *one* input light that had to be *on* in order for TM to turn green when it was pressed, otherwise TM turned red. In HEPP II problems there was *one* input light that had to be *off* in order for TM to turn green when it was pressed, otherwise TM turned red. There were two reasons for using both types of problem in spite of their similarity: there is some evidence that schizophrenics have special difficulty in controlling set (Shakow, 1963); the inverted relation between HEPP states (on-off) and the information signal (red-green) occurs frequently in the more complex problems of the series, and it was, therefore, important to obtain an indication of the effect of the inversion at this relatively simple

FIG. 1. *S*'s panel.

level. Transfer from one class to the other would be expected to be positive and complete, except for disturbance associated with the reversal of set. In this study, however, it will not be possible to obtain a satisfactory measure of this effect, because the various problem categories were given to all Ss in the same sequence.

Test Sequence

Using roman numerals to indicate problem class and arabic numerals to indicate the number of lights in the display field, problems were presented in the following order: I-8, I-16, II-8, II-16. Each S was given five problems in each of these categories. Except to inform S of the change, there was no interruption between problems of a given class. The instructions for the Class II problems were given immediately after completion of the Class I quota, and testing was resumed. In most cases, all problems were completed in one session that varied from 30 to 50 min. depending upon S's rate of work.

RESULTS

It will be recalled that the instructions required S to locate the HEPP with as few inquiries (TM presses) as possible. He was free to adopt whatever strategy appeared to him to be appropriate to this end. The operating characteristics of his plan can be displayed quite simply by plotting the number of possibilities that remained after each trial or selection. Data presented in this manner in Figure 2 were selected to illustrate some fundamental points. The curve for S 756 shows the operating characteristic of the optimal plan. On each trial S tested half of the remaining possibilities and, thus, with this strategy, would always solve the 16 light problems in exactly four trials. The S 712 tried lights in sets of various sizes until one that contained HEPP was encountered and, then, he proceeded with an item-by-item test on the members of that set until HEPP was identified. This procedure does not show as much insight as the optimal strategy, but it is superior to the item-by-item test illustrated by S 627, who was lucky enough to encounter HEPP on his fifth trial. Through the operation of a random success factor, this and the previous procedure can occasionally have quite happy outcomes, success in one and two trials, respectively, but they also can be especially hazardous, if not applied in an orderly way. The dangers of disorderly search are illustrated in the data from S 773. He

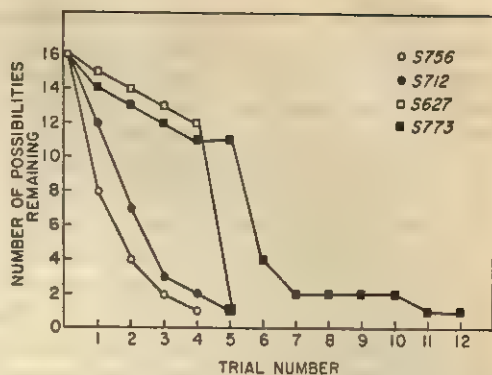


FIG. 2. Selected possibility functions.

used a mixture of the last two plans with no apparent pattern in his selections and made several redundant choices (Trials 5, 8, 9, and 10) that suggest serious disorientation.

These possibility functions provide pictures that are useful for reviewing the performances of individual patients, but they are not appropriate for group comparisons. Some kind of data reduction, with its attendant loss of detail, is necessary in the interest of comprehensibility. Since the amount of work done by each S was fixed, the groups might be compared with respect to the average number of trials required for solution. Associated with this measure, however, is the random success factor mentioned in connection with S 627's fifth trial. This effect would greatly increase the error variance and render tests of significance insensitive. Moreover, such a score would not reflect the strategy employed with a satisfactory degree of accuracy. The index described below, though somewhat complicated, does overcome these difficulties.

Expected Information Value (EIV)

The HEPP problems were designed to permit an evaluation of each selection made by S in the course of his solution effort. In fact, there are two such measures for each trial: (a) the amount of information actually elicited by the selection and (b) the amount of information that it might rationally be expected to elicit. The former is subject to the random success factor, but the latter is not affected by the outcome at all. Expected information value (EIV) depends only upon

the selection in relation to the uncertainty structure in which it is made. Thus, at the beginning of any trial, there are N equally probable solutions, and the existing uncertainty is $\log_2 N$. The display chosen by S for testing partitions the possibility set into two subsets, one consisting of N_1 possibilities that remain if the display is classified in one way, say green, the other consisting of $N - N_1 = N_2$ possibilities that remain if it is classified in the other way, say red. The corresponding uncertainties are $\log_2 N_1$ and $\log_2 N_2$, representing information gains of $\log_2 N - \log_2 N_1$ and $\log_2 N - \log_2 N_2$, respectively. The EIV of a selection is the average of these two possible gains weighted for their separate probabilities of occurrence. Since the original N possibilities are assumed equiprobable, the probability of occurrence of either gain depends only upon the number of possibilities included in the associated subset; that is, the probability of a gain of the first magnitude is N_1/N and that of the second is N_2/N . The expected information value of the selection, thus becomes $EIV = N_1/N (\log_2 N - \log_2 N_1) + N_2/N (\log_2 N - \log_2 N_2)$ which, for convenient reference to existing tables, (e.g., Attneave, 1959) can be rewritten $EIV = 1/N (N \log_2 N - N_1 \log_2 N_1 - N_2 \log_2 N_2)$. The EIV ranges from 0 (no information obtained) to 1 (half of the existing possibilities eliminated). Throughout the remainder of this report, indexes referred to as "problem scores" will be averages obtained by weighting each trial EIV by the number of possibilities existing just prior to the selection to which it refers. In the tables, these indexes have been multiplied by 100 to eliminate decimal points.

Relative Efficiency

Figure 3 summarizes the general outcome of the experiment and shows the average problem score of each group for the various problem classes and sizes. It rather strongly suggests that the performances of the schizophrenics were inferior to those of the controls. For each of the four problem categories separately, a two-way analysis of variance with repeated measures was performed. The only reliable effect was that due to problem order

in the I-8 and I-16 categories ($p < .05$). Similarly, t tests on the difference between means for individual problems failed to reveal any in which the groups differed significantly. In both of these tests, the error variance was determined by the between-Ss variability confounded with other random factors. Since schizophrenics are known to exhibit abnormally large individual differences in a variety of performances (Shakow, 1963), such tests are likely to be insensitive and to fail to detect real differences between groups. However, it will be noticed that the controls had the better EIV on 15 of the 20 problems. Each pair of problem means provides an estimate of the relative proficiency of the two groups. When the distribution of differences generated by these paired comparisons was used to estimate an error variance, the unimpressive mean difference of 1.7 in favor of the controls was found to be statistically significant ($p < .02$).^a

The general upward trend from early to late problems, seen in several segments of Figure 3, indicates that both group performances improved with experience. The reliability of the observed gains from the first to the fifth problem within each category was estimated by determining whether the mean of the 15 within-Ss differences was significantly greater than zero. On this criterion, the controls were found to have made reliable gains ($p < .03$) in all problem categories except II-16 and the schizophrenics improved reliably in I-8 ($p < .01$) and II-16 ($p < .02$) only, despite large apparent gains in both I-16 and II-8.

Changes in problem category were associated with two types of cognitive demand. Those from I-8 to I-16 and from II-8 to II-16 involved increments in uncertainty; that from I-16 to II-8 involved a reversal of the information cue and a decrease in uncertainty. As a result of the progressive complexity design of the study, these two latter effects, which would be expected to influence performance in opposite directions, can not be separated or evaluated independently. However, since both groups performed at

^a One-tailed p values have been used consistently for the results of t tests.

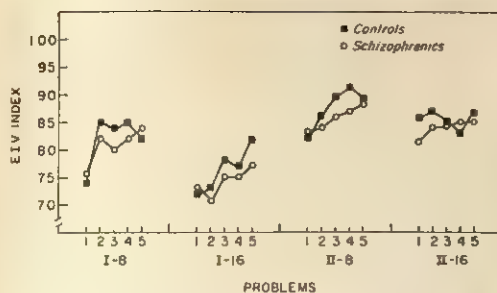


FIG. 3. Average EIVs for schizophrenic and control groups on successive problems.

least as well on the first problem of II-8 as they had on any previous problems, there was no direct evidence of disturbance associated with the reversal of set. It may be noted, however, that the schizophrenics showed a large gain, though nonsignificant ($p < .16$), from the last problem of I-16 to the first of II-8, whereas the controls had a small, but unreliable, loss at this point. Thus, if the schizophrenics were disturbed by the reversal, as their comments frequently suggested they were, the reduction in problem size enabled them to more than compensate for losses associated with the inversion. If the controls derived any benefit from the reduction in uncertainty, it was completely masked by the effect of information reversal. That changes in uncertainty had considerable influence was shown by (a) a significant loss ($p < .04$) in both groups associated with the shift from I-8 to I-16, an increment in uncertainty; (b) the schizophrenics' gain at the switch from I-16 to II-8, when uncertainty was decreased; and (c) the schizophrenics' significant loss at the shift from II-8 to II-16 ($p < .03$), when complexity was increased again. Since the performances of the controls were not significantly affected at these latter transition points, the evidence suggests that the schizophrenics were more sensitive to uncertainty than the controls.

Subtypes of Schizophrenia

Average HEPP scores according to problem category and type of schizophrenic patient included in the experimental group are shown in Table 2, which exhibits a marked contrast between the performances of para-

TABLE 2
AVERAGE EIVs BY SUBCLASSIFICATIONS
OF SCHIZOPHRENIA

Subclassifications	N	I-8	I-16	II-8	II-16
Chronic undifferentiated	9	89	84	91	87
Chronic paranoid	4	69	58	75	85
Schizo-affective	1	57	55	66	48
Schizoid personality	1	83	67	95	88

noid and chronic undifferentiated schizophrenics. A more detailed comparison appears in Figure 4, which, except for the separation of schizophrenic groups, is analogous to Figure 3. Again using differences between pairs of problem means to estimate the error variances, it was found that the difference in favor of the undifferentiated schizophrenics over the controls was significant at the .01 level. The differences between chronic undifferentiated and paranoid schizophrenics, and between controls and paranoids were significant at the same level.

Figure 4 also provides an explanation for the failure of the method of gains to detect a reliable improvement in I-16 and II-8 for the schizophrenics despite a rather large apparent gain from the first to the fifth problem in these categories. It will be noticed that, whereas the four paranoids made large gains on the average, the undifferentiated schizophrenics apparently had achieved a rather high plateau, and consequently made little or no average gain. This bimodality, of course, made for a large error variance and an asso-

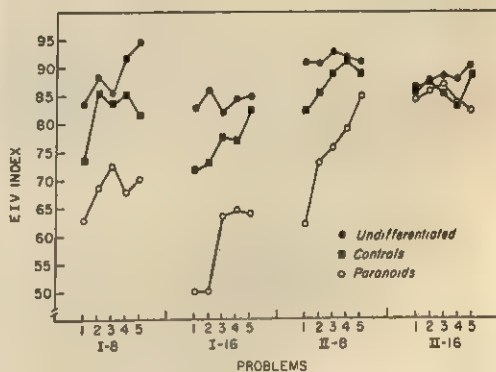


FIG. 4. Average EIVs for chronic undifferentiated and paranoid schizophrenics and control groups on successive problems.

ciated lack of sensitivity. This, in turn, was a result of differences in practice effects among the two subtypes of patients.

Correlations

The HEPP index employed in computing the rank-order correlation coefficients presented in the following tables was obtained by averaging the five problem scores of each *S* within the relevant problem categories. Table 3 gives the correlations between performances on the various types of problems with the two *S* groups combined. All correlations are positive and significant at less than the .01 level. Their magnitudes indicate a high interclass reliability of HEPP performances. Moreover, since the observed trend in amount of correlation is in the direction of a priori similarity, it is suggested that the within-class reliability is probably even higher. These results also indicate that the considerable between-category improvement of performance with experience, observable in Figure 3, was not accompanied by a marked rearrangement of the relative proficiencies of individual *Ss*, at least when the combined groups are considered.

Comparable indexes are given for the two groups separately in Table 4. It will be observed that the schizophrenics contributed more than the controls to the overall correlations; for every pair of correlates, their coefficient was higher than that of the controls. Moreover, for the schizophrenics, all correlations were significant whereas, for the controls, only intraclass correlations were reliable. The difference in correlation pattern between the two groups indicates that the relative proficiencies of individuals was more stable among the schizophrenics than it was among the controls. This, in turn, may be

TABLE 3

CORRELATIONS BETWEEN HEPP PERFORMANCES,
GROUPS COMBINED

HEPP	I-16	II-8	II-16
I-8	.73*	.50	.47
I-16		.54	.67
II-8			.69

* $p < .01$.

TABLE 4

CORRELATIONS BETWEEN HEPP PERFORMANCES

HEPP	Schizophrenics			Controls		
	I-16	II-8	II-16	I-16	II-8	II-16
I-8	.82**	.58*	.61*	.63**	.28	.39
I-16		.63**	.77**		.42	.62*
II-8			.77**			.66**

* $p < .05$.

** $p < .01$.

attributable to the marked separation between paranoids and undifferentiated schizophrenics noted in connection with Figure 4.

The correlations between HEPP performances and the extraexperimental variables reflected by WAIS performance, education, and age, are displayed in Table 5 for the combined group. HEPP II performances clearly correlated more frequently with these variables than did HEPP I performances. This suggests that some experience with HEPP-like problems was required before the cognitive processes associated with the extraexperimental variables manifested their influence significantly. Age was the only variable listed that did not correlate significantly with any HEPP performance in this combined group.

In Table 6, comparable indexes for the two groups are listed separately. There were no significant correlations with HEPP I scores among the controls, but the schizophrenics did have a reliable correlation between verbal intelligence and HEPP I-16. However, these variables did not manifest their general relevance to the schizophrenics' performance until the HEPP II-16 problems were given, whereas it was manifest among the controls for the HEPP II-8 problems, that is, somewhat earlier. Finally, it should be noted that the controls showed a significant relation between HEPP II-16 and age. The failure of the schizophrenics to do so may be a result of the restricted age range in this group (15-40 yr.) compared with the controls (16-65 yr.). Moreover, since all except one of the correlations between HEPP performances and age were negative, it appears quite possible that there is a real relation between performance and age that was

not detectable under the conditions of this experiment.

DISCUSSION

The results of this study indicate that the procedure employed elicited from schizophrenics performances on elementary HEPP problems that were at least as efficient as those of control Ss. The implications of this outcome depend, of course, upon the nature of the problem involved and upon the possibilities for future developments that this success entails. A HEPP problem of any degree of difficulty confronts *S* with the task of planning and executing an inquiry into the operation of an artifact that behaves in a lawful way. The *S* may be given much, or little, information about an artifact that operates according to simple, or complex, principles, and, depending on these variables, he may be required to adduce, by exploration and reasoning, much, or little, about the laws that are controlling this test segment of reality. The HEPP situation requires that *S* execute a sequence of meaningful, goal-directed transactions, or interactions, with the artifact. Because he controls the information flow, his actions may be minimally or maximally effective; he may create ambiguities and difficulties for himself by a poorly designed plan of inquiry, or he may minimize the cognitive strain of the solution effort by sagacious planning. Finally, because he can be exposed repeatedly to formally identical situations that differ only with respect to the particulars of the solution, he can be given numerous opportunities to criticize his pro-

TABLE 5
WAIS SCORES, EDUCATION, AND AGE CORRELATED
WITH HEPP PERFORMANCES, GROUPS COMBINED

Item	I-8	I-16	II-8	II-16
WAIS				
Verbal	.11	.34*	.41*	.43**
Performance	.33*	.29	.40*	.34*
Full scale	.20	.33*	.49**	.49**
Education	.06	.21	.36*	.63**
Age	-.19	-.21	-.30	-.25

* $p < .05$.

** $p < .01$.

cedure, correct its faults, and thus to profit by experience.

The present study indicates that schizophrenics are capable of performing these tasks effectively; hence, that they are capable of understanding the artifact-testing assignment and have intact those cognitive and heuristic processes necessary for executing it with "normal" efficiency at the elementary level. Moreover, since schizophrenics manifested elation on success, chagrin at failure, and improvement with experience, without the use of noxious reinforcement, it appears that these gamelike problems have, for them, intrinsic motivating components that many other tests fail to exhibit. Given the basic ability and motivation shown in this study, schizophrenics should be able to work on increasingly difficult HEPP problems until they encounter some specific cognitive demand to which they are not prepared to respond effectively.

The marked difference between the performances of paranoid and chronic undiffer-

TABLE 6
WAIS SCORES, EDUCATION, AND AGE CORRELATED WITH HEPP PERFORMANCES

Item	Schizophrenics				Controls			
	I-8	I-16	II-8	II-16	I-8	I-16	II-8	II-16
WAIS								
Verbal	.28	.47*	.43	.74**	-.09	.10	.40	.24
Performance	.28	.31	.33	.44*	.40	.21	.48*	.25
Full scale	.20	.36	.41	.64**	.10	.11	.63**	.38
Education	.22	.29	.14	.57*	-.12	.00	.71**	.58*
Age	-.10	-.09	-.34	.19	-.27	-.37	-.26	-.51*

* $p < .05$.

** $p < .01$.

entiated schizophrenics, noted in connection with Figure 4, is regarded as a far from well-founded finding. Although, on the average, these groups did not differ markedly with respect to duration of hospitalization, 5.5 yr. and 6.4 yr., nor WAIS intelligence, 105.5 and 106, for the undifferentiated and paranoid, respectively, the latter were, on the average, 9 yr. older than the former. Since, however, HEPP performances did not correlate significantly with age among the schizophrenics, it is unlikely that a large performance disparity would be associated with this factor. Nonetheless, before hypothesizing a specific heuristic difference between paranoid and chronic undifferentiated schizophrenics, it is highly desirable to expand both groups in the interest of obtaining possibly more representative samples.

REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
- ARIETI, S. Schizophrenia: The manifest symptomology, the psychodynamic and formal mechanisms. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1959.
- ATTNEAVE, F. *Applications of information theory to psychology*. New York: Holt, 1959.
- BLATT, S. J. Patterns of cardiac arousal during complex mental activity. *Journal of Abnormal and Social Psychology*, 1961, 63, 272-282.
- DEWEY, J. *Logic: The theory of inquiry*. New York: Holt, 1938.
- GOTTESMAN, I. I., & CHAPMAN, L. J. Syllogistic reasoning errors in schizophrenia. *Journal of Consulting Psychology*, 1960, 24, 250-255.
- JEROME, E. A. Decay of heuristic processes in the aged. In C. Tibbitts & W. Donahue (Eds.), *Social and psychological aspects of aging*. New York: Columbia University Press, 1962.
- JEROME, E. A., YOUNG, M. L., & UMBERGER, S. A problem programmer. *Perceptual and Motor Skills*, 1964, 19, 711-717.
- JOHN, E. R. Contributions to the study of the problem-solving process. *Psychological Monographs*, 1957, 71(18, Whole No. 447).
- MILLER, G. A., GALANTER, E., & PRIEBRAM, K. H. *Plans and structure of behavior*. New York: Holt, 1960.
- NIMS, J. P. Logical reasoning in schizophrenia: The von Domarus principle. Unpublished doctoral dissertation, University of Southern California, 1959.
- RIMOLDI, H. J. A., & DEVANE, J. R. *Training in problem solving*. Loyola Psychometric Laboratory, 1961. (Mimeo)
- SHAKOW, D. Psychological deficit in schizophrenia. *Behavioral Science*, 1963, 8, 275-305.
- VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: University of California Press, 1946.
- WEINER, I. B. *Psychodiagnosis in schizophrenia*. New York: Wiley, 1966.
- WILLIAMS, E. B. Deductive reasoning in schizophrenia. *Journal of Abnormal and Social Psychology*, 1964, 69, 47-61.
- YOUNG, M. L. Problem-solving performance in two age groups. *Journal of Gerontology*, 1966, 21, 505-509.

(Received August 31, 1967)

CLIENT-COUNSELOR COMPATIBILITY AND THE OUTCOME OF COUNSELING¹

GERALD A. MENDELSON² AND NEIL O. RANKIN³

University of California, Berkeley

Client-counselor compatibility as measured by the Fundamental Interpersonal Relations Orientation Behavior (FIRO-B) scale, was related to clients' ($N = 115$) perceptions of the relationship and evaluations of the counselor and the usefulness of counseling. Compatibility was a poor predictor for male clients but an excellent one for females. Generally, compatibility in the control need area was related positively to outcome, but surprisingly, compatibility in the inclusion and affection need areas was related to negative outcomes. Consequently, the global compatibility measure used successfully by Sapolsky yielded no significant correlations in this study. However, a composite formed a posteriori for the females produced a correlation of .54 with the major outcome variable. These results point to the necessity of examining potential sex differences, particularly in studies of the clinical relationship, and to the possibility of empirically based client-counselor matching.

In a recent paper, Sapolsky (1965) reported that "the degree of interpersonal compatibility [in terms of Schutz's, 1958, Fundamental Interpersonal Relations Orientation Behavior, FIRO-B, scale] existing between patient and doctor was a significant variable affecting the outcome of hospital treatment [p. 75]." Since patients in high compatibility pairs felt more similar to and better understood by their doctors, and feelings of being understood were related to favorable outcomes, he suggests that the relationship between compatibility and outcome is mediated by the effect of compatibility on the patient's perception of his doctor. Although he notes the small size of his sample and the need for more data, he concludes that compatibility may be an important variable underlying the establishment of "good" therapeutic relationships.

It is surprising that FIRO-B has been so little used in clinical research, for it is virtually unique in providing a direct operational

measure of interpersonal compatibility; the test is specifically designed for use in studies of dyadic and group behavior. Sapolsky's results are encouraging as a validation of a potentially useful test, as an empirical confirmation of ideas about the importance of the clinical relationship to outcome, and as an indication of the possibility of rational patient-doctor matching. There are, however, a number of limitations of the study which raise serious questions about the generality and meaning of his findings. Not only is the sample small in size, but it also consists entirely of females treated in a hospital by three first- and second-year residents. Further, there are several methodological questions involved in the use of compatibility scores which are overlooked in the study. It is the intent of the present investigation to examine the effects of compatibility in a very different clinical setting and by a quite different approach to analysis. If compatibility in terms of FIRO-B is really an "important underlying personality variable contributing to the establishment of 'good' therapeutic relationships [Sapolsky, 1965, p. 75]," one would expect comparable (although not identical) results to those of Sapolsky irrespective of the setting and the approach to analysis. It should be noted at the outset that the study to be reported is not a replication of

¹ This research was supported by Vocational Rehabilitation Administration Grant RD-1741-P-66. The authors wish to express their thanks to the staff of the Counseling Center of the University of California, Berkeley, and especially to the center's Director, Barbara A. Kirk, for advice, assistance, and cooperation.

² Requests for reprints should be sent to Gerald A. Mendelson, Department of Psychology, University of California, Berkeley, California.

³ Now at Clark University.

Sapolsky's work but rather is complementary to it.⁴

METHOD

Subjects and Data Collection Procedure

The Ss of this study were 162 clients, 104 males and 58 females, who had come to the Counseling Center of the University of California, Berkeley, seeking assistance with vocational, educational, and personal problems. Every client coming to the Counseling Center for the first time during the data collection period was asked to participate in a research project designed to improve counseling services. Clients were told that this would entail about an hour of testing and that the results of the tests would not be available to their counselors but would be reserved for research purposes only. More than 90% of the clients asked to participate agreed to do so; lack of time was the usual reason given for refusal. Use of the Counseling Center's services is voluntary, so that neither the decision to come to the Center nor to participate in the study was the result of coercion or administrative decision.

The clients were seen by 11 counselors; 4 of them (two males and two females) had 5 yr. or more of counseling experience, 4 (two males and two females) had between 2 and 5 yr. of experience, and the remaining two males and one female were graduate students in clinical psychology who had less than a year of counseling experience. The counselors in each experience level saw an approximately equal number of clients. Three-quarters of the clients were assigned to a counselor of their own sex. The counseling was of short duration, eight interviews being the maximum in this sample. The orientation of counseling is toward viewing the client's problems within a developmental and personality framework as well as in terms of the current environment. The client-counselor relationship is explicitly emphasized as a critical factor in the counseling. Consequently, although the setting and the Ss of the present research are different from those studied by Sapolsky, the clinical relationship cannot be considered any less important to the counseling process than to other clinical processes.

Prior to the beginning of the data collection, all participating counselors took FIRO-B and two other self-report inventories not germane to the present paper. Each client took the same set of tests before the first interview. After the conclusion of the series of interviews, an evaluation questionnaire was mailed to each client in the sample. A series was considered to be concluded when 3 mo. had elapsed after the last recorded interview. If a client did not return the questionnaire within 2 wk., a follow-up letter with another questionnaire enclosed was sent. A total of 115 clients, 73 males and 42 females, 71% of the initial sample, returned usable questionnaires.

Independent Variables

A full description of the development and theoretical rationale of FIRO-B and the procedure for deriving compatibility scores from it are given in Schutz (1958). The test provides scores in three need areas, inclusion (I), control (C), and affection (A) which Schutz argues constitute a sufficient set of dimensions to predict interpersonal behavior. The test attempts to measure both the extent to which the *S* expresses behavior toward others in each area and the extent to which he wants others to express the behavior toward him. Thus each *S* received six scores: expressed inclusion (*I_e*), wanted inclusion (*I_w*), etc. Schutz then delineates three kinds of compatibility, reciprocal compatibility (*rK_{ij}*), originator compatibility (*oK_{ij}*), and interchange compatibility (*xK_{ij}*); the subscript *i* refers to one member of a dyad and the subscript *j* to the other member. The meaning and specific formula for each is given below.

Reciprocal compatibility "reflects the degree to which members of a dyad reciprocally satisfy each others behavior preferences [Schutz, 1958, p. 108]." The formula is: $rK_{ij} = |e_i - w_j| + |e_j - w_i|$. A zero score reflects maximum compatibility and the larger the score, the less the compatibility.

Originator compatibility is concerned with the balance of initiating and receiving behavior in the dyad. Compatibility is present when the two members have complementary patterns.

Conflict arises when there is disagreement regarding preference of who shall originate relations and who shall receive them. For each need area (I, C, A) there are two types of conflict: between two originators, *competitive originator incompatibility*, and between two receivers, *apathetic originator incompatibility* [Schutz, 1958, p. 109].

The formula is: $oK_{ij} = (e_i - w_i) + (e_j - w_j)$. Positive scores reflect competitive incompatibility and negative scores apathetic incompatibility. Again a score of zero indicates maximum compatibility.

Interchange compatibility "refers to the mutual expression of the 'commodity' of a given need area [Schutz, 1958, p. 110]." It can be seen from the formula that it is, in effect, a measure of the similarity or dissimilarity in behavior preferences between dyad members: $xK_{ij} = |(e_i + w_i) - (e_j + w_j)|$. As with reciprocal compatibility, the larger the sum the less the compatibility. It is true in all cases that zero indicates maximum compatibility.

Since there are three test dimensions and three forms of compatibility, there are nine compatibility scores which will be referred to as "individual compatibility scores." In addition, Schutz suggests a number of composite scores, of which only one is of present concern: the sum of all nine scores, designated *K*. It is this measure which was used by Sapolsky. The authors will refer to it in this paper as the "global compatibility score."

There are several problems inherent in the use of matching scores of any kind in counseling re-

⁴The data for the study were collected prior to the appearance of Sapolsky's paper.

search (see Mendelsohn, 1966) and particularly in the use of composite measures of matching. The most important problem is that matching scores are necessarily a function of the separate scores from which they are derived and may thus, be correlated with them to a greater or lesser degree. Consequently, what one attempts to explain by matching may, in fact, be the result of client or counselor scores alone. If one examines the formula for originator compatibility, for example, it is apparent that it is nothing more than an unweighted arithmetic combination of two client and two counselor scores; any results attributed to originator compatibility can equally well be attributed to the client and counselor scores. This relationship between the component scores and the compatibility scores does not hold for the measures of reciprocal or interchange compatibility, however, because both those measures involve absolute difference scores. While they may be related to client and counselor scores, unlike the measure of originator compatibility, they include an independent matching component. Since Sapolsky does not examine the separate effects of the client and counselor scores, it is difficult to reach any conclusions about the independent effects of compatibility in his study.⁵

Two additional problems arise in the use of the unweighted sum of the nine compatibility scores, that is, K , as a measure of global compatibility. Both reciprocal and interchange compatibility are absolute scores, but originator compatibility can have both positive and negative values. Thus when the nine scores are summed, K will be larger if there is "competitive originator incompatibility" and smaller if there is "apathetic originator incompatibility." This leads to a situation in which as one form of incompatibility increases, overall compatibility, as operationally defined, also increases. The problem could be handled if originator compatibility scores were treated as absolute values, but neither Schutz nor Sapolsky makes any such suggestion. Again this leads to difficulty in interpreting Sapolsky's findings in terms of compatibility. The second problem in the use of a global measure like K is that its components may contribute differentially to the prediction of a criterion. A single component or a subset of components may predict as well or better than K and it is even possible that two components may predict the criterion in opposite directions. In the latter case, a simple sum in which all components are weighted equally can lead to a cancellation of effects. Under any circumstances, a global measure is more difficult to interpret than its individual components and a differentially weighted composite is likely to predict better than a composite formed of equally weighted components.

⁵ Sapolsky is not, it seems, unaware of this problem but the attempt to control for it (p. 73) does not provide information on separate effects of patient or doctor scores.

The authors have gone into some detail in the discussion of FIRO-B and the measures of compatibility derived from it both to point out certain questions about Sapolsky's findings and to indicate the methodological problems involved in the use of matching scores. The authors will necessarily return to these problems in the data analyses to be reported.

Outcome Questionnaire

The outcome questionnaire used in this study is the same as that described in Mendelsohn and Geller (1965). It consists of 27 items, each in the form of a 7-point Likert-type scale. Clients are asked to evaluate a number of aspects of their counseling experience and to indicate their reactions to the counselor and the process. Cluster analyses of client responses in four separate samples have produced a relatively invariant structure which applies equally well to males and females. It consists of five clusters whose alpha reliabilities are .85, .80, .77, .59, and .78, respectively: Cluster 1 represents degree of general satisfaction with counseling and evaluation of the extent to which objectives were achieved; Cluster 2 taps evaluation of the counselor's skill and perceptiveness; Cluster 3 reflects the client's feelings of comfort and of being understood in counseling; and the last two clusters are concerned with specific ways in which counseling was of benefit, through the therapeutic means of insight and catharsis (Cluster 4) and by acquiring concrete information and reaching specific decisions about future actions (Cluster 5). Though the five obtained clusters consist of different sets of items and refer to conceptually different aspects of the counseling, their intercorrelations are sufficiently high to suggest that there is a single dimension of evaluation of the counselor and the counseling which underlies much of the variance of the questionnaire responses. Consequently, the first centroid was extracted (Thurstone, 1947) and the 11 items with the highest factor loadings on it were selected to define the centroid. This set of items had an alpha reliability of .94 and accounts for 61% of the initial item communality. There are thus six dependent measures available, but since the first centroid has such generality, it will be used as the major outcome variable in the presentation of results. It will be referred to as "general evaluation."

Scores were obtained for the clusters and the centroid by summing a client's ratings on the defining items and converting the resulting distributions to T scores. Items were scored in such a way that the higher the T scores, the more favorable the client's response.

The outcome measures of the present study are different from those used by Sapolsky. In his study, evaluation of outcome was by the supervising psychiatrist and measures of relationship and rapport were inferred from the profile similarity of patient and doctor Semantic Differential protocols. In contrast, the client is the source of the outcome vari-

ables in the present study and the measures of relationship were obtained by direct question. Although self-report has its limitations as a source of evaluation of success, it remains the least equivocal method of assessing the client's feelings about and perceptions of the counselor and the process of counseling.

It is, then, only in the use of the same predictor, FIRO-B, that the two studies are the same, but as noted before, if compatibility is a central factor in fostering "good" relationships, results of the two studies should be comparable.

RESULTS

The correlations between the 10 compatibility scores and the measure of general evaluation (first centroid) are shown for males and females in Table 1.

Perhaps the most striking aspect of these results is the difference between the correlations for males and females. For the latter, 5 of the compatibility scores correlate significantly ($p < .05$) with general evaluation, but for the former none of the correlations even approaches significance. This basic finding holds for the correlations between the compatibility scores and the five other clusters scores as well. There is at least one significant predictor for each cluster score for the females. In contrast, only two significant ($p < .05$) correlations were obtained for males and both are relatively low ($r = .25$). Since 60 correlation coefficients were

computed for each sex, the conclusion that the results for males are due to chance cannot be rejected. Thus evidence of a relationship between FIRO-B compatibility and outcome was obtained for females only.^a

Before reaching any conclusions about the relationship of compatibility scores to the criteria in the female sample, the possibility that client or counselor scores alone can predict as well as or even better than the compatibility scores must be examined. The findings for the general evaluation score are typical and will serve to illustrate the pattern of results. Two client scores, Wanted Control (C_w) and Expressed Control (C_e) have moderate correlations with the criterion ($r = .28$ and $-.27$, respectively), but neither is significant at the .05 level and neither is as high a correlation as any of the five significant ones between compatibility scores and general evaluation (see Table 1). Compatibility, then, does seem to have an important predictive function, independent of client or counselor scores, for the females.

The second finding of note in these data is the failure of the global compatibility measure (K), used by Sapolsky, to predict any of the outcome measures for either sex significantly. For females, this failure occurs despite the presence of a substantial number of significant correlations between the outcome measures and the individual compatibility scores which are the components of K . Inspection of Table 1 reveals the reason for this failure quite clearly. As noted before, the global compatibility measure is the simple sum of the individual compatibility scores, but these nine scores do not all correlate with the criterion in the same direction. In fact, for each sex, just about half the correlation coefficients are positive and half are negative. Consequently, the effects of the individual compatibility scores cancel each other out when these scores are combined into the global compatibility measure. However, these

^a The same analyses were performed for groups composed of same-sex and opposite-sex client-counselor pairs. The results were generally quite similar for these two groups and the small differences between them which were found were attributable to the differences, reported above, between male and female clients.

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN COMBATIBILITY SCORES AND GENERAL EVALUATION

Need area		Males ^b	Females ^c
Inclusion			
Reciprocal	(rK^I)	.09	.35*
Originator	(oK^I)	.10	-.17
Interchange	(xK^I)	.05	.30*
Control			
Reciprocal	(rK^C)	-.12	.09
Originator	(oK^C)	-.02	-.35*
Interchange	(xK^C)	-.11	-.32*
Affection			
Reciprocal	(rK^A)	-.12	.34*
Originator	(oK^A)	.01	-.06
Interchange	(xK^A)	-.11	.25
Global compatibility	(K) ^a	-.05	.13

^a Global Compatibility is the sum of the nine component compatibility scores listed above it.
^b $n = 73$.
^c $n = 42$.
^{*} $p < .05$.

variations in direction are not random—every significant correlation between a compatibility score in the I and A need areas and an outcome measure is positive in sign and all the significant correlations involving the C need area are negative in sign. Since maximum compatibility is defined as a score of zero and the higher the outcome score, the more favorable the client's evaluation, these results indicate that compatibility in I and A need areas is associated with less client satisfaction, lower evaluation of the counselor, and less comfort and rapport on the part of the client. The expected relationship between compatibility and outcome was found only in the C need area.

Although K , the global compatibility measure suggested by Schutz, proved of no value as a predictor in this study, it remains possible that some other composite of the nine individual compatibility scores could produce higher correlations with the criteria than does any single compatibility term. The authors have noted two difficulties with the K measure, one methodological and the other empirical, which should be avoided in the formation of such a composite. The methodological problem concerns the measure of originator compatibility. Unlike the other compatibility measures, it can take on both negative and positive values and is not really a compatibility measure at all. Thus originator compatibility terms should be excluded in the formation of a composite. The second difficulty is that in the formation of K all terms are weighted equally, that is, they are all given a weight of +1. However, it was found that some compatibility scores correlate positively and some negatively with the criteria and thus tend to cancel each other when combined. In order to avoid this problem, the terms entering the composite should be weighted in the direction of their correlation with the criteria. If a measure of overall compatibility has any utility at all, a composite formed in this way should yield appreciably higher correlations than those based on the individual compatibility terms. The prediction of general evaluation scores in the female sample was selected as a test case.

The composite score was obtained in the following manner. First all individual com-

TABLE 2

PRODUCT-MOMENT AND MULTIPLE CORRELATION COEFFICIENTS BASED ON FIRO-B COMPOSITE SCORES (FEMALES FOR GENERAL EVALUATION)

Predictor	Correlation
Compatibility composite	.50***
Client-counselor composite	.39*
Multiple regression	.54**

* $p < .05$.

** $p < .01$.

*** $p < .001$.

patibility scores were standardized. Then each compatibility score which was significantly ($p < .05$) correlated with the criterion was given a weight of +1 when the correlation was positive and -1 when the correlation was negative. Thus, the composite consisted of $rK^I + rK^A + xK^I - xK^C$ and a single composite score was derived for each client by summing his weighted standard scores. If, for example, a client's standard scores on the four measures were 45, 55, 40, and 35, respectively, his composite score would be 105. For purposes of comparison, a composite based on the significant ($p < .10$) client and counselor score predictors was formed in an identical fashion.⁷ These two composites were then correlated with the general evaluation score and, in addition, a multiple-regression coefficient using both composites as predictors was obtained. The results are shown in Table 2.

Perhaps the clearest way to evaluate the utility of the compatibility composite is to compare the proportion of variance explained by the composite to the proportion of variance explained by the best individual compatibility score predictor, that is, by a comparison of r^2 . The best single predictor is reciprocal compatibility in the I need area—it accounts for 12% of the variance of the criterion. In comparison, the compatibility

⁷ Composites can also be found on the basis of multiple-regression analysis, but the procedure reported is (a) simpler, (b) more similar to the manner in which K , to which it is compared, is formed, and (c) likely to be more stable than a multiple-regression composite. Nevertheless multiple correlation coefficients were also calculated for these data; they are virtually identical to those reported in Table 2.

score composite accounts for 25% of the variance. Clearly the use of the composite does result in an appreciable improvement in prediction, doubling the amount of explained variance. This stands in sharp contrast to the results obtained for the *K* composite, but this finding should be interpreted cautiously for the following reasons. The composite was formed a posteriori in an attempt to maximize prediction. Thus, while the data show that a composite *can* improve prediction, cross-validation is necessary to evaluate the stability of this particular composite. Further, some terms enter the composite with positive and some with negative weights; consequently, despite its practical value, it is difficult to ascribe a meaning, particularly in terms of compatibility, to the composite.

The authors noted before that there is a problem in making the inference that results are attributable to client-counselor matching (compatibility) rather than to the characteristics of the client or counselor alone. The data of Table 2 also provide information on this point. If compatibility does have a predictive function independent of the client and counselor scores, the multiple-regression prediction based on both composites should be significantly better than the prediction based on the client-counselor composite alone. A comparison of the two coefficients (.39 versus .54) indicates that this is the case. In contrast, there is little difference between the multiple-regression correlation and the correlation based on the compatibility score composite alone (.50 versus .54). Furthermore, the compatibility score composite remains a significant predictor of outcome after the effect of the client-counselor score composite is partialled out ($r_{12.3} = .41$, $p = .01$). Thus, the evidence indicates that FIRO-B compatibility scores are important independent predictors of outcome for the female client.

DISCUSSION

This study was undertaken to examine the hypothesis that client-counselor compatibility, as measured by FIRO-B, would be related to the outcome of counseling. Such a finding would, of course, add strength to

Sapolsky's suggestion that compatibility may contribute to the establishment of a good therapeutic relationship. Although the authors' results, like those of Sapolsky, indicate that some scores derived from FIRO-B are effective predictors of outcome, the difference between the results of the two studies are considerable. Perhaps the most clear-cut difference is that the global compatibility measure, *K*, which works quite well in Sapolsky's study, is a failure in the present study. The effects of compatibility in different need areas, it seems, can be directly opposite, a possibility obscured by the use of the *K* score. While it should be kept in mind that the two studies differ radically in the samples and processes studied, the evidence casts serious doubt on the generality of the positive relationship between "compatibility" and favorable outcomes found by Sapolsky.

Further doubt on the generality of this relationship arises from the differences in results for males and females. Since Sapolsky studied only females, no comparisons are possible, but it is clear that for the present sample, compatibility, as operationalized by Schutz, is a variable of consequence only for females. This may reflect the well-substantiated finding that, compared to males, females are more sensitive and responsive to interpersonal behavior and are more dependent on others (Tyler, 1965). Consequently, compatibility is more likely to affect their response to counseling than it is to affect the response of the males. But whether or not this interpretation of the obtained sex difference is correct, it is clear that generalizing about the clinical relationship or the clinical process on the basis of results for one sex is a dangerous procedure. Yet the number of studies in this area which make any attempt to explore potential sex differences is quite small. Given the consistency with which differences in social behavior between males and females have been found, this seems an unfortunate omission which may as easily lead to the rejection of valid but limited hypotheses as to the overgeneralization of findings.

The significant correlations between compatibility scores and outcome variables follow a consistent, but surprising pattern: compatibility in the I and A need areas is

related to unfavorable outcomes and compatibility on the C dimension is related to favorable outcomes. If the FIRO-B compatibility indexes really measure what they are supposed to measure, the results for C are expectable and support the notion that the direction of the counseling process should be shared by the two participants. The results for the I and A need areas, the two dimensions most concerned with the affective aspects of relationships, on the other hand, appear paradoxical. However, findings of other studies of the effects of client-counselor matching (Carson & Heine, 1962; Mendelsohn, 1966; Mendelsohn & Geller, 1965, 1967) suggest that conditions which encourage closeness can have a deleterious effect on the clinical process. Carson and Heine, for example, suggest that high similarity of personality can lead on the part of the therapist to an overidentification with the patient and his problems. It seems a tenable hypothesis that factors which foster strong emotional attachments in typical social relations can, unless carefully handled, lead to an excessive personalization of the special relationship which characterizes counseling and psychotherapy.

Without comparable data from other clinical settings, it is impossible to know whether the specific results found in the present study will be found elsewhere. There are, moreover, other, simpler, compatibility scores possible which could be as effective, or more effective, than those suggested by Schutz. Despite these reservations and despite the differences between the present study and that of Sapolsky, one central point emerges—FIRO-B scores can generate some remarkably good predictions of outcome, at least for females. Considering that (a) the tests are administered before the beginning of counseling and neither participant knows his own or his partner's test results, (b) the results attributed

to compatibility scores cannot be explained in terms of client or counselors scores alone, (c) there are differences among counselors in background and style and among clients in objectives and maturity, (d) there is a myriad of important events which take place in the counseling itself, and (e) there is a considerable time lapse between the testing and evaluation of outcomes, a correlation of the magnitude of .54 can be described without exaggeration as remarkable. That such a result was obtained suggests the potential importance and utility of systematic, empirically based, matching of client to counselor. Effective matching alone can hardly guarantee success, but the results of this study and of Sapolsky's strongly suggest that it is a feasible and practicable way to facilitate favorable clinical outcomes.

REFERENCES

- CARSON, R. C., & HEINE, R. W. Similarity and success in therapeutic dyads. *Journal of Consulting Psychology*, 1962, 26, 38-44.
- MENDELSON, G. A. Effects of client personality and client-counselor similarity on the duration of counseling: A replication and extension. *Journal of Counseling Psychology*, 1966, 13, 228-234.
- MENDELSON, G. A., & GELLER, M. H. Structure of client attitudes toward counseling and their relation to client-counselor similarity. *Journal of Consulting Psychology*, 1965, 29, 63-72.
- MENDELSON, G. A., & GELLER, M. H. Similarity, missed sessions, and early termination. *Journal of Counseling Psychology*, 1967, 14, 210-215.
- SAPOLSKY, A. Relationship between patient-doctor compatibility, mutual perception, and outcome of treatment. *Journal of Abnormal Psychology*, 1965, 70, 70-76.
- SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Holt, Rinehart & Winston, 1958.
- TEURSTONE, L. L. *Multiple-factor analysis*. Chicago: University of Chicago Press, 1947.
- TYLER, L. E. *The psychology of individual differences*. (3rd ed.) New York: Appleton-Century-Crofts, 1965.

(Received August 25, 1967)

PSYCHOTHERAPEUTIC CHANGE AND SOCIAL ADJUSTMENT:

A REPORT OF A NATIONAL SURVEY OF PSYCHOTHERAPISTS¹

RUTH K. GOLDMAN AND GERALD A. MENDELSON²

University of California, Berkeley

A representative national sample of psychotherapists (psychiatric social workers, psychologists, and psychiatrists) was asked to describe by means of Gough's Adjective Check List (a) themselves, (b) the type of patient with whom they work best (preferred), (c) a patient who has received maximum benefit from treatment (cured), and (d) a person with satisfactory adaptation to himself and his environment (normal). Despite the heterogeneity of the sample, there is a remarkably high level of agreement among therapists in their descriptions. Therapists feel they work best with a patient who exhibits little pathology and the successfully treated patient is described almost entirely in terms of positive social functioning. Therapists do not see the preferred, cured, or normal as similar to themselves. These data suggest an implicit consensus about the goals of psychotherapy and point to the salience of a social adjustment criterion in the evaluation of outcome.

During the past two decades, there has been a growing recognition of the contribution of the psychotherapist to the treatment process. Strupp (1960) describes this trend as a "rediscovery" of the therapist as a person, whose attitudes, personality and values presumably play an exceedingly important, if not crucial, part in the therapeutic process [pp. 66-67]. Undoubtedly, one important influence on the values and attitudes, if not the personality, of the individual therapist is the set of beliefs held by his fellow practitioners, first as they are his teachers and then as they are his peers. Yet, one knows very little about what therapists as a group are like or what they believe about psychotherapy. Usually studies of therapists have used small samples, comprised of Ss drawn from a single clinic or a limited geographical area and confined to only one profession, most often psy-

chology. The authors know of only one study to date in which a national sample of therapists of all relevant professions was used (Chance & Arnold, 1960), but in that study, all the participants came from six major cities.

The present research was undertaken to obtain normative information which could make possible some generalization about therapists' self-perceptions, and their values and attitudes toward the patients they treat. The approach was to ask a representative, nationwide, sample of therapists to describe themselves, the patient at the outset and at the termination of therapy, and the "normal" person. These descriptions were then used to draw inferences about some of the conceptualizations therapists share about who can profit from therapy and the nature of psychological well-being.

METHOD

Procedure

Twelve hundred psychotherapists were randomly selected from the national directories of the American Psychiatric Association (1961-62), the Clinical Division of the American Psychological Association (1962), and the Psychiatric Social Work section of the National Association of Social Worker's Directory (1960). Materials consisting of four adjective checklists, a Personal History Questionnaire (PHQ), detailed instructions, and a cover letter were sent to this group of therapists. At the end of a 2-mo. period, a second set of materials was

¹ This research was supported by Public Health Service Grant 4F1-MH-12, 595-04, from the National Institute of Mental Health, and is based in part on a doctoral dissertation submitted to the University of California at Berkeley in 1964. A substantial portion of the work on this paper was done at the Institute of Personality Assessment and Research, University of California, Berkeley. The authors wish to express their particular appreciation to Harrison G. Gough and Quintin Welch of the institute for their advice and assistance.

² Requests for reprints should be sent to Gerald A. Mendelsohn, Department of Psychology, University of California, Berkeley, California 94720.

sent to all nonrespondents. Three months after the mailing of the second letter, all returns were analyzed; materials which arrived after this date were not included in this analysis.

Instruments

Therapists were asked to describe the following four concepts by means of Gough's Adjective Check List: (a) "check those adjectives which describe you" (self), (b) "check those adjectives which describe the kind of adult male patient that you work with best" (preferred), (c) "check those adjectives which describe an adult male patient that has terminated therapy successfully" (cured), and (d) "check those adjectives which describe an adult male who has a satisfactory adaptation to himself and his environment" (normal). Order of presentation was balanced in order to avoid response set due to sequence.

Gough's Adjective Check List (ACL) is a standardized list of 300 words which provides a particularly efficient way to obtain a wide range of behavioral descriptions of people. A detailed presentation of the rationale, development, and standardization of the checklist is available in Gough and Heilbrun (1965). For the authors' purposes, it has the following particular merits: It utilizes no special language—in particular it is free of psychiatric and theoretical terminology which might act to restrict the range of behavior described; it is not oriented toward pathology; it is easily understood; it can be scored objectively; and it is consistent with everyday modes of describing people. The value of the ACL as a descriptive instrument has been discussed by Gough (1960) and is well illustrated by MacKinnon's (1963) study of creative architects.

The Personal History Questionnaire was devised in order to obtain information about each therapist's background and training. They were asked to indicate the following: profession, sex, religion, parents' socioeconomic status, years of personal psychotherapy, years of experience as a psychotherapist, and theoretical persuasion. These data were used in the present study to evaluate the representativeness of the sample of respondents. In subsequent papers, the effect of these variables on the descriptions of the concepts will be reported.

Response Rate

A pilot study was conducted in order to determine an expected response rate. The results of this study indicated that the return rate of psychologists and psychiatric social workers was twice that of psychiatrists. Therefore, in the final study, twice as many psychiatrists as social workers or psychologists were contacted in order to obtain approximately equal numbers of respondents in each profession.

Of 1,200 therapists to whom the materials were sent, 186 could not be reached, 443 responded, and of this number, 421 provided data suitable for an-

alysis. By profession 29% ($n = 144$) of psychiatrists, 61% ($n = 143$) of clinical psychologists, and 60% ($n = 134$) of psychiatric social workers responded. These figures represent about 5% of the total population in each profession. The response rate of psychiatrists is consistent with that generally obtained for mail surveys, but for the other two professions, it is considerably higher than one would expect (Moser, 1958; Sundland & Barker, 1962; Wallach & Strupp, 1960).

Sample Characteristics

The characteristics of the sample appear in Table 1.

Since the intent of this study was to obtain normative data about psychotherapists, it is important to ascertain whether respondents are a representative sample of the population from which they were drawn. Two surveys of mental health specialists are available for evaluation of this question: the Department of Health, Education, and Welfare (HEW) studies on mental health manpower (1963) and the American Psychiatric Association manpower study (1963a, 1963b). A minimum of 80% of the members of the national psychiatric, psychological, and social work organizations participated in the HEW

TABLE 1
SAMPLE CHARACTERISTICS

Characteristic	Percentage
Sex	
Male	71
Female	29
Religion	
Protestant	45
Catholic	12
Jewish	40
None	3
Socioeconomic status	
Upper	32
Upper middle	14
Middle middle	14
Lower middle	20
Lower	20
Personal Psychotherapy	
None	42
Under 3 yr.	24
Over 3 yr.	34
Experience	
Under 5 yr.	16
5-10 yr.	46
Over 10 yr.	38
Profession	
Psychiatrists	34
Psychologists	34
Social workers	32
Theoretical Persuasion	
Psychoanalytic	63
Interpersonal	25
Eclectic	12

Note.— $N = 421$.

survey and the American Psychiatric Association provides data on its total membership.³

The samples of psychiatrists and psychiatric social workers are virtually identical in age, sex, and type of clinical employment to the populations from which they were drawn. In addition, the data on years of experience and religious affiliation of psychiatrists in the present sample were strikingly similar to those reported in the HEW study. Thus, these data and those of the Department of Health, Education, and Welfare are quite similar for psychiatrists and social workers, although there is the least amount of comparison data available for the latter group.

The sample of clinical psychologists is somewhat less representative of the national population surveyed by HEW. However, the discrepancies between the data for this sample and those for the HEW group are small; the differences, when they appear, are of the order of 10%. The present sample of psychologists contains more females, self-employed therapists, and residents of California than the HEW group does. The data on age and years of experience, on the other hand, are very similar. Considering these comparisons for all three professions, there is reason to have confidence in generalizing from the results obtained for the authors' sample to the population of psychotherapists as a whole.

RESULTS AND DISCUSSION

The results will be presented in two parts: in the first, the adjectives most frequently chosen by the therapists to describe each concept will be considered and in the second, the focus will be on the adjectives which differentiate between concepts.⁴

Self-Concept

The 25 adjectives most frequently checked to describe the self are: dependable (90%), capable (86%), conscientious (86%), intelligent (86%), friendly (85%), honest (85%), adaptable (83%), responsible (82%), reasonable (81%), reliable (81%), considerate (80%), cooperative (80%), understanding

(80%), active (79%), sincere (79%), alert (78%), realistic (78%), fair-minded (76%), healthy (75%), interests-wide (75%), sensitive (74%), curious (73%), helpful (73%), tolerant (73%), and warm (72%).

It is apparent from inspection of this list that the proportion of therapists endorsing these adjectives is very high. This finding is not limited to the self-description, but is characteristic of the results for all four concepts. This may not seem surprising since these adjectives were selected for presentation because they were the most frequently checked. But it is important to keep in mind that there are 300 adjectives to choose from and that on the average, a therapist checked 100 of them; that is, the chance probability of any given adjective being checked is .33. Furthermore, the sample is quite heterogeneous with respect to profession, experience, training, sex, socioeconomic origins, and theoretical inclination. Consequently, these percentages are much higher than expected; there is reason to have confidence in characterizing the attitudes of psychotherapists in general from these data.

The self-portrait which emerges is of a person who is steady, competent and intellectually alert, and concerned with the welfare and feelings of others. Adjectives associated with originality, maturity, enthusiasm, and spontaneity are checked less frequently however. This contrast is even more striking when the responses of the psychotherapists are compared to those of architects (MacKinnon, 1963). At least 80% of both groups describe themselves as capable, conscientious, cooperative, friendly, honest, intelligent, and reasonable. The therapists describe themselves more frequently than the architects as adaptable, considerate, dependable, reliable, responsible, and understanding. The architects describe themselves with considerably greater frequency than the therapists as imaginative, idealistic, civilized, and serious. Both groups, then, consider themselves worthy of trust, skilled, and sensitive to other people. The therapists place greater emphasis on consistency and dependability. The most striking difference between the two groups though is the architects' greater con-

³ For clinical psychologists, there were no significant differences between respondents and nonrespondents in the Health, Education, and Welfare study.

⁴ Gough and Heilbrun provide a set of 24 scales for scoring the ACL, but these scales were standardized for self-description only. Consequently, their applicability to the descriptions of the "preferred," "cured," and "normal" is questionable. It seemed to the authors, too, that item analyses would provide the clearest, most direct, and most detailed picture of the therapists' descriptions.

cern for originality and their greater sense of personal involvement.

This pattern of results leads one to conclude that the therapists were describing themselves, in part, in the role of therapist and, in part, as individuals; to the extent that this is true, the high level of agreement among therapists points to a considerable consensus on how the therapist ought to behave toward patients. Stability, objectivity, and a willingness to help are stressed but, as could be expected, indications of affect are considerably less emphasized.

Preferred Patient

The 25 adjectives most frequently checked to describe the preferred patient are: intelligent (75%), alert (74%), adaptable (70%), frank (68%), honest (68%), cooperative (67%), anxious (65%), sincere (65%), sensitive (64%), curious (62%), active (61%), interests-wide (61%), conscientious (60%), imaginative (58%), friendly (57%), insightful (56%), dependable (55%), capable (54%), reasonable (54%), reflective (54%), clear thinking (53%), thoughtful (53%), warm (53%), appreciative (51%), and responsible (50%).

There is less agreement among therapists in their description of the preferred patient than there is about any of the other concepts. Nevertheless, the agreement is sufficiently high to provide a consistent description of the kind of patient with whom the therapists work best. He is an imaginative, sensitive, curious, well-motivated, but anxious person. Aside from the anxiety, there is little indication of pathology or a "cry for help" on his part. Adjectives such as despondent, high-strung, hostile, impulsive, suspicious, timid, unstable, and withdrawn, to cite several examples, are checked by fewer than one-quarter of the therapists. In fact, if anxious were removed from the list of the 25 most descriptive adjectives, the preferred patient would appear to be an unusually productive and creative person.

There are several possible interpretations of what at first seems a puzzling finding. The obtained pattern suggests that therapists have found in practice trends which parallel

those of a number of empirical studies, namely, that patients who benefit most from psychotherapy are those with the least amount of manifest pathology (Barron, 1953; Sullivan, Miller, & Smelser, 1958; Taulbee, 1958). This finding raises serious issues when one considers that the patient described by therapists as most capable of profiting from treatment is not typical of the majority of those who feel in need of psychological help (Gurin, Veroff, & Feld, 1960).

Cured Patient

The 25 adjectives most frequently checked to describe the cured patient are: adaptable (86%), dependable (79%), honest (77%), alert (76%), reasonable (76%), confident (75%), friendly (75%), realistic (75%), responsible (75%), active (72%), clear thinking (72%), capable (71%), self-confident (71%), cooperative (70%), considerate (69%), sincere (69%), insightful (68%), tolerant (67%), reliable (66%), mature (65%), rational (65%), stable (65%), natural (64%), understanding (64%), and appreciative (63%).

Agreement among therapists in describing a patient who has terminated therapy successfully is again quite high. The cured patient, as described by therapists, is a person who is very well adapted to his environment. There are no indications of psychological disturbance; instead he has acquired self-confidence, contentment, and a measure of tolerance and stability. He is, above all, a good citizen, capable of taking a responsible role in society. But except for the adjective natural there is little indication that he is seen as a freer, more creative, more unique individual. The criterion of successful termination seems to be based on social effectiveness rather than on self-actualization.

Normal Person

The 25 adjectives most frequently checked to describe the normal are: adaptable (94%), dependable (90%), capable (87%), confident (86%), friendly (86%), honest (85%), realistic (85%), clear-thinking (84%), alert (83%), cooperative (83%), reasonable (83%), responsible (83%), active

(82%), considerate (81%), mature (80%), sincere (79%), tolerant (79%), affectionate (78%), appreciative (78%), reliable (78%), self-confident (78%), warm (78%), conscientious (77%), stable (76%), and tactful (76%).

The highest level of agreement among therapists is found in their description of the normal person. There is a marked bimodality in the distribution of the percentages for the adjectives—either they are checked by more than a majority of therapists or they are checked very infrequently, that is, by less than 10% of the Ss.

An "adult male who has a satisfactory adaptation to himself and his environment" is seen as possessing all the positive, but none of the negative characteristics of the other three concepts. He is reality oriented, socially effective, self-accepting, free of inner conflict, and concerned with the welfare of others. But in contrast to the self, preferred, and cured, the normal is described as more affectively involved in his relationships. It is the one description which gives a sense of balance between emotion and control.

Comparisons between Concepts

In the discussion which follows, the emphasis will be on those adjectives which differentiate between concepts to the greatest extent.

Self-Concept Compared with Preferred Patient Concept

Two-thirds of the adjectives differentiate between the two concepts at the .05 level of confidence or beyond.⁵ In order to highlight the differences between the self and the preferred, the 12 items with the greatest discrimination in favor of the self and the 12 with the greatest discrimination in favor of the preferred patient will be presented. In this and in subsequent comparisons of con-

cepts, only those adjectives for which a difference of 10% or more between concepts was obtained will be included.

The 12 adjectives which differentiate most in favor of the self (versus preferred) are: helpful (73% versus 27%), understanding (80% versus 38%), healthy (75% versus 34%), tactful (71% versus 30%), considerate (80% versus 41%), sympathetic (72% versus 33%), loyal (65% versus 28%), discreet (56% versus 20%), forgiving (65% versus 29%), reliable (81% versus 45%), dependable (90% versus 55%), and independent (65% versus 30%).

The 12 adjectives which differentiate most in favor of the preferred (versus self) are: dissatisfied (45% versus 16%), confused (32% versus 3%), fearful (37% versus 9%), anxious (65% versus 38%), immature (31% versus 10%), despondent (23% versus 5%), nervous (33% versus 16%), awkward (22% versus 8%), self-punishing (26% versus 13%), unstable (14% versus 1%), withdrawn (19% versus 7%), and distrustful (17% versus 6%).

There are two primary dimensions on which the concepts are differentiated. The first concerns interpersonal relations and the second psychopathology. Compared to the preferred patient, the therapists describe themselves as compassionate and responsive to the needs and feelings of other people. The preferred patient is described expectably as more disturbed, self-absorbed, and subject to inner tension than the self. The discrimination is considerably greater on the first dimension than on the second, however; consider, for example, the adjective anxious, the pathological trait most characteristic of the preferred. It is checked by 65% of therapists for the preferred and by 38% of therapists for themselves. The percentage difference of 27% is less than that for any of the dozen adjectives discriminating in favor of the self. Thus, it seems that the self is distinguished from the preferred more by positive qualities than by the negative ones associated with pathology. It will be seen that this is not characteristic of the other two comparisons in which the preferred appears.

⁵ Since the same therapists provide data on all four concepts, it is necessary to take this potential source of dependency in the data into account in performing statistical tests for differences between concepts. This was done by using the chi-square test for a repeated measurements situation suggested by McNemar (1949, pp. 204-207). It is based on who changes response from one concept to the next.

Self-Concept Compared with Cured Person Concept

More than two-thirds of the adjectives differentiate between the two concepts at the .05 level of confidence or beyond.

The 12 adjectives which differentiate most in favor of the self (versus cured) are: intelligent (86% versus 43%), sentimental (49% versus 13%), serious (69% versus 33%), idealistic (57% versus 25%), reserved (42% versus 14%), soft-hearted (34% versus 6%), quiet (40% versus 13%), conscientious (86% versus 60%), curious (73% versus 48%), cautious (49% versus 24%), impatient (28% versus 4%), and anxious (38% versus 15%).

There are only five adjectives which differentiate in favor of the cured (versus self) and meet the additional criterion of a 10% difference in percentage checked. They are self-confident (72% versus 50%), confident (75% versus 55%), relaxed (56% versus 37%), assertive (63% versus 51%), and contented (43% versus 32%).

Perhaps the most striking difference between the two concepts is the greater self-assurance and confidence ascribed to the cured. In contrast, the self is described as more introverted and less at ease with the world. But the therapists also see themselves as more intellectually open, thoughtful, and affectively responsive than the cured. There is a sense of vitality in the self-description which is lacking in the description of the cured.

Self-Concept Compared with the Normal Person Concept

Three-fifths of the adjectives differentiate the self-concept from that of the normal person at the .05 level of confidence or beyond.

The 12 adjectives which differentiate most in favor of the self (versus normal) are: sentimental (49% versus 20%), anxious (38% versus 10%), serious (69% versus 43%), reserved (42% versus 16%), impatient (28% versus 2%), cautious (49% versus 24%), tense (27% versus 3%), complicated (37% versus 15%), argumentative

(26% versus 4%), emotional (38% versus 16%), inhibited (24% versus 4%), and stubborn (24% versus 4%).

The 12 adjectives which differentiate most in favor of the normal (versus self) are: confident (86% versus 55%), self-confident (78% versus 50%), relaxed (64% versus 37%), wholesome (60% versus 37%), courageous (59% versus 36%), contented (53% versus 32%), calm (70% versus 49%), mature (80% versus 64%), natural (72% versus 56%), spontaneous (62% versus 46%), resourceful (71% versus 56%), and assertive (66% versus 51%).

The tendency of the therapists to describe themselves as prone to conflict and lacking in self-assurance is most evident in the comparison of the self and normal. This tendency to include a moderate level of negative characteristics in the self-description appeared also, but to a lesser degree, in the comparison of the self and cured. However, the greater sense of vitality which differentiates the self from the cured does not differentiate the self from the normal.

Preferred Patient Concept Compared with Cured Patient Concept

Three-quarters of the adjectives differentiate the preferred from the cured patient at the .05 level of confidence or beyond.

The 12 adjectives which differentiate most in favor of the preferred (versus cured) are: anxious (65% versus 15%), dissatisfied (45% versus 6%), intelligent (75% versus 43%), fearful (37% versus 6%), confused (32% versus 3%), talkative (40% versus 12%), nervous (33% versus 6%), defensive (33% versus 7%), emotional (47% versus 21%), immature (31% versus 5%), inhibited (32% versus 6%), and outspoken (42% versus 16%).

The 12 adjectives which differentiate most in favor of the cured (versus preferred) are: confident (75% versus 29%), self-confident (71% versus 30%), stable (65% versus 31%), contented (43% versus 9%), mature (65% versus 33%), independent (62% versus 30%), realistic (76% versus 45%), relaxed (56% versus 25%), self-controlled (53% versus 23%), affectionate (62% ver-

sus 34%), tactful (57% versus 29%), and healthy (60% versus 33%).

As one would expect, the adjectives which discriminate between the preferred and cured patient concepts reflect psychopathology on the one hand and psychological well-being on the other. The three adjectives which deviate from this generalization are talkative, intelligent, and outspoken. The significance of the first needs no comment but the presence of the latter two is of considerable interest. The authors noted in the discussion of the adjectives most characteristic of the preferred patient, the appearance of an important dimension of intellectual vigor, sensitivity, interest, openness, indeed of creativity. If one examines differences between the preferred and cured patient concepts on this dimension, one finds the preferred described more often than the cured as artistic ($p < .001$), clever ($p < .001$), complicated ($p < .001$), spunky ($p < .001$), curious ($p < .01$), idealistic ($p < .02$), original ($p < .05$), imaginative ($p = .05$), and individualistic ($p = .05$). These data suggest that while the cured person is described as having achieved maturity, self-control, stability, and contentment, he does not appear self-actualized or creative, or a genital character, etc. Anxiety, defensiveness, and socially maladaptive behavior are absent compared to the preferred, but so too are originality and zest.

Preferred Patient Concept Compared with Normal Person Concept

Four-fifths of the adjectives significantly differentiate the preferred and normal person concepts at the .05 level of confidence or beyond.

The 12 adjectives which differentiate most in favor of the preferred (versus normal) are: anxious (65% versus 9%), dissatisfied (45% versus 3%), fearful (37% versus 1%), confused (32% versus 1%), nervous (33% versus 2%), emotional (47% versus 17%), immature (31% versus 1%), tense (32% versus 3%), talkative (40% versus 12%), inhibited (32% versus 4%), defensive (33% versus 5%), and worrying (29% versus 3%).

The 12 adjectives which differentiate most in favor of the normal (versus preferred) are:

confident (86% versus 29%), self-confident (78% versus 30%), mature (80% versus 33%), tactful (76% versus 30%), calm (71% versus 25%), stable (76% versus 31%), contented (53% versus 9%), affectionate (78% versus 35%), independent (71% versus 30%), helpful (68% versus 27%), realistic (85% versus 45%), and considerate (81% versus 41%).

The adjectives which discriminate most between the preferred and the normal are largely the same as those which discriminate most between the preferred and cured. However, those adjectives descriptive of creativity and intellectual vitality, for example, curious, imaginative, original, etc., are no less characteristic of the normal than of the preferred. The differentiation of the two concepts, then, is purely in terms of the contrast between psychopathology and social psychological well-being.

Cured Patient Concept Compared with Normal Person Concept

The cured and the normal person concepts are the most similar of all the pairs, but still one-half of the adjectives significantly differentiate the two. The only adjectives which discriminate significantly in favor of the cured are endorsed on the average by only 6 or 7% of the therapists and consequently do not fulfill the criterion for inclusion in the comparison, that is, a 10% difference between the two concepts in percentage checked. They refer exclusively to small residues of pathological traits in the cured, for example, fearful, immature, etc. Since this is the case, the authors will present the 20 adjectives which differentiate most in favor of the normal. They are intelligent (66% versus 43%), good-natured (57% versus 35%), generous (61% versus 41%), loyal (57% versus 37%), wholesome (60% versus 40%), courageous (59% versus 40%), humorous (61% versus 43%), tactful (76% versus 58%), ambitious (70% versus 52%), civilized (65% versus 47%), enthusiastic (63% versus 45%), progressive (55% versus 37%), thorough (49% versus 31%), adventurous (50% versus 32%), wise (42% versus 24%), conscientious (77% versus

60%), discreet (56% versus 39%), versatile (51% versus 34%), original (45% versus 28%), and resourceful (71% versus 55%).

In examining this list, it is as important to note which adjectives do not appear as to note those which do. There is no important difference between the descriptions of the two concepts with regard to stability, competence, and reliability. Both the cured and the normal are seen as well adjusted in a socially defined sense. But the normal is also seen as possessing to a greater degree characteristics which transcend adjustment. Many theorists speak of psychotherapy as an experience which frees human potentialities and fosters the growth of a unique identity. It is surprising, then, to find that it is a person who has "a satisfactory adaptation to himself and his environment" who more nearly approaches this ideal than a person who "has terminated therapy successfully."

CONCLUSIONS

The conclusions which emerge from this study are, first, that there is a marked consensus among therapists about, at least, some important aspects of psychotherapy. The authors do not mean to assert that all therapists have the same conceptions about all aspects of therapy, but rather that there are some attitudes and conceptions which can be described as characteristic of psychotherapists in general. This conclusion is in contrast to the impression gained from such similar studies as those of Strupp (1958) and Sundland and Barker (1962). It likewise contrasts with the opinion of Colby (1964) that, "In the domain of psychotherapy . . . there is no simple shared paradigm commanding consensus."

It seems to the authors that whether one concludes that diversity prevails or that a consensus prevails is dependent upon what question is asked. If the concern is with technique and the questions are phrased in language which is drawn from particular theoretical orientations, it is likely that a picture of differences among therapists will emerge (Strupp, 1958; Sundland & Barker, 1962). However, if, as in the present case, a theoretically neutral language is used and the

emphasis is on goals, rather than means to goals, a picture of consensus emerges. The extent of agreement about the description of a person who has a satisfactory adaptation to himself and to his environment moreover, suggests that it is possible to derive an empirical criterion of "positive mental health," at least, as this elusive concept is conceived by specialists in the field.

Another point which emerges with considerable clarity concerns the relationship of the therapist to the patient he treats. There is little evidence either that the therapist feels that he works best with those patients who most resemble him, or that he sees the patient who has obtained maximum benefit from treatment as closely resembling himself. For that matter he does not see himself as normal (as he has defined it). This suggests that the therapist does not use himself as the point of reference in establishing treatment goals, but rather adheres to an external criterion defined, it seems, by the society in which he and his patients reside.

The final point to be noted is the modesty of the expectations therapists hold for psychotherapy. This can be seen with respect to both the patient when he enters therapy, and when he has successfully terminated therapy. As pointed out earlier, the patient with whom therapists say they work best is intelligent, alert, well motivated, reliable, and lacking in serious psychopathology; anxiety is the only major indication of psychological distress. In parallel fashion, the description of the successfully treated patient stresses his self-control and competence, his adaptation to social expectations.

There is little of the self-actualized person in this description; rather it is adjustment that seems to be the main emphasis. While those who criticize psychotherapy for fostering conformity may find support for their view in these data, the authors would argue that they reflect what therapists believe to be practicable, rather than ideal, and point to a realistic appraisal of what therapy can achieve. Osgood and Luria (1954) in their study of a case of multiple personality pose the same question raised by these data, "In striving to achieve the goals of societal ac-

ceptability and individual happiness, does the therapist have to sacrifice the richness, individuality and subtler adjustiveness of the patient?" Neither their study nor this one can provide an answer, but it is a question which is too serious to ignore.

REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION. *List of fellows and members of the American Psychiatric Association*. New York: American Psychiatric Association, 1961-62.
- AMERICAN PSYCHIATRIC ASSOCIATION, Ad Hoc Committee on Manpower. *Psychiatric Manpower Bulletin*. Washington, D. C.: American Psychiatric Association, April 1963. (a)
- AMERICAN PSYCHIATRIC ASSOCIATION, Ad Hoc Committee on Manpower. *Psychiatric Manpower Bulletin*. Washington, D. C.: American Psychiatric Association, May 1963. (b)
- AMERICAN PSYCHOLOGICAL ASSOCIATION. *1962 Directory*. Washington, D. C.: American Psychological Association, 1962.
- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology*, 1953, 17, 327-333.
- CHANCE, E., & ARNOLD, J. The effect of professional training, experience and preference for a theoretical system upon clinical case description. *Human Relations*, 1960, 13, 195-213.
- COLBY, K. Psychotherapeutic processes. *Annual Review of Psychology*, 1964, 15, 347-370.
- GOUGH, H. G. The adjective check list as a personality assessment research technique. *Psychological Reports*, 1960, 6, 107-122.
- GOUGH, H. G., & HEILBRUN, A. B. *Manual for the Adjective Check List*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.
- GURIN, G., VEROFF, J., & FELD, S. *Americans view their mental health*. New York: Basic Books, 1960.
- MACKINNON, D. W. Creativity and images of the self. In R. W. White (Ed.), *The study of lives*. New York: Atherton, 1963.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
- MOSER, C. A. *Survey methods in social investigation*. London: Heinemann, 1958.
- NATIONAL ASSOCIATION OF SOCIAL WORKERS. *Directory of Professional Social Workers*. New York: National Association of Social Workers, 1960.
- OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *Journal of Abnormal and Social Psychology*, 1954, 49, 579-591.
- STRUPP, H. The psychotherapist's contribution to the treatment process. *Behavioral Science*, 1958, 3, 34-67.
- STRUPP, H. Some comments on the future of research in psychotherapy. *Behavioral Science*, 1960, 5, 60-71.
- SULLIVAN, P. L., MILLER, C., & SMELSER, W. Factors in length of stay and progress in psychotherapy. *Journal of Consulting Psychology*, 1958, 22, 1-9.
- SUNDLAND, D. M., & BARKER, E. N. The orientations of psychotherapists. *Journal of Consulting Psychology*, 1962, 26, 201-212.
- TAULBEE, E. S. Relationship between certain personality variables and continuation in psychotherapy. *Journal of Consulting Psychology*, 1958, 22, 83-89.
- UNITED STATES DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, NATIONAL INSTITUTE OF MENTAL HEALTH. *A Mental Health Manpower Studies Program*. Bethesda, Md.: National Institute of Mental Health, March 1963.
- WALLACH, M. S., & STRUPP, H. Psychotherapists' clinical judgments and attitudes toward patients. *Journal of Consulting Psychology*, 1960, 24, 316-323.

(Received September 4, 1967)

CONSISTENCY OF SELF-DESCRIPTIONS UNDER DIFFERENT ROLE SETS IN NEUROTIC AND NORMAL ADOLESCENTS AND ADULTS¹

DAVID G. MARTIN²

University of Iowa

Fifteen male Ss in each of four groups (neurotic adolescents, normal adolescents, neurotic adults, and normal adults) were asked to describe themselves with a Q sort as they saw themselves in three different situations (self in general, self with family, and self with friends). The tendency to describe themselves inconsistently was characteristic of neurotic adolescents specifically and not of adolescents or of neurotics in general. These results were discussed in relation with adolescence as a time of "identity crisis" and with previously reported findings that neurotics tend to have inconsistent self-concepts.

The importance of self-consistency to personality has probably been stressed most strongly by Lecky (1945) although many others have embodied the notion in their writing. Lecky stated as a motive the tendency to create a unified personality which the individual defines consistently for himself and others. Other theorists who have stressed the development of a consistent "self" include Kelly (1955), Allport (1955), and Rogers (1951).

As Wiley (1961) has pointed out in her summary of the literature on the self-concept, self-consistency has been postulated as descriptive of personality but little attention has been paid to the antecedents which lead to this consistency. In addition, most self-theorists have treated the self as a construct or an entity. In contrast with this, it is the author's position that the tendency toward self-consistency (defined here as consistency of self-description) is an acquired behavior and therefore partly a developmental problem. The present investigation was based on two lines of previous research: the relationship between self-concept consistency and the psychology of adolescence and between this consistency and emotional disturbance.

¹This study was supported by the Institute for Juvenile Research, Chicago, Illinois, and is based on part of a study done as partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Chicago.

²Requests for reprints should be sent to the author, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

Erickson (1956) has defined three criteria of "ego identity": (a) the adolescent's perception of himself as essentially the same person over time and interpersonal situations, (b) others' perceptions of the adolescent as essentially the same over time and situations, and (c) an accrued confidence on the part of the adolescent that his perception of himself corresponds to others' perceptions of him. This is a more inclusive definition than has been used in previous research, which has been focused on the first of these criteria. It is a widely held notion that adolescence is a time of "identity crisis" for all young people (Bronson, 1959; Erikson, 1956; Heilbrun, 1964; Hess & Goldblatt, 1957; Hood, 1962; Howard, 1960; Musgrove, 1964; Rosenberg, 1965). However, none of the literature cited has been based on a comparison between adolescents and adults. Further, Engle (1959) reported "relative stability of the self-concept" among adolescents tested over a 2-yr. period, supporting her position that "crystalization of the self-concept is achieved earlier in development." Dividing her data into disturbed and normal groups, Engle also found that the disturbed group had significantly less consistency in their self-concepts than had the normal group.

A second issue is the relationship between emotional disturbance and self-consistency. The available evidence seems to indicate that emotionally disturbed people have greater flux in their self-concepts than do normal people (Block, 1961; Brownfain, 1952; Butler, 1956; Cartwright, 1961; Engle, 1959;

Gough & Heilbrun, 1965; Hickman, 1959; Phillipsson, 1955; Tolor, 1955; Woldstad, 1961). However, with the exception of one (Gough & Heilbrun, 1965),³ all of the studies reviewed in connection with this problem used undergraduate Ss and were actually examinations of the relationship between self-concept inconsistency and emotional disturbance in late adolescence.

The above observations suggested that the tendency to describe oneself inconsistently may be a prominent concomitant of emotional disturbance in adolescence, but that it is less frequently a characteristic of the disturbed adult; the presence of emotional disturbance may delay the successful clarification of a stable self-description. It was predicted that (a) neurotic adolescent boys would describe themselves less consistently under different role sets than would normal adolescent boys and (b) neurotic boys would differ from normal boys on this measure to a greater extent than neurotic men would differ from normal men.

METHOD

Instruments, Data Collection, and Scoring

The instrument used was the Butler-Haigh *Q* sort (1954), modified to ensure readability of the items for a young group. Each *S* completed the *Q* sort three times. First Ss were asked to describe themselves the way they were "in general," arranging the cards in a normal distribution of nine piles. The Ss were then asked to describe "yourself the way you are with your friends" and "yourself the way you are when you are with your family." The second and third sortings were reversed in order of presentation for half of the Ss.

The operational definition of consistency of self-description was the mean squared correlation of the three possible combinations of *Q* sorts (r_{12} , r_{13} , r_{23}).

Judgments of the "intensity of emotional disturbance" were made for each of the neurotic Ss on a 5-point scale (1—none, 2—mild, 3—moderate, 4—marked, 5—very intense). A formal psychiatric diagnosis was also made by staff conferences at an outpatient clinic to which each of the disturbed Ss had applied for services.

³ Gough and Heilbrun, moreover, did not report results related specifically to emotional disturbance. Their data showed that men with lower test-retest reliability on the ACL were described by observers as having more unpleasant qualities.

Subjects

A factorial design was used, and four groups of 15 Ss each were tested: neurotic adolescent boys, normal boys, neurotic adult males, and normal adult males. All Ss were Caucasian and, among the adolescents, had both parents living at home. The neurotic adolescents had an average IQ of 106. Previous large-scale testing programs done at the school from which the normal adolescent Ss were selected had resulted in IQ averages of 105–112 for different classes. All Ss were from middle-class homes, following Warner (1960). The adolescents ranged from 13 to 16 yr. old. The disturbed adolescent sample included all boys who fit the criteria and who had made their first contact with the clinic in the preceding 4 mo. The disturbed adult sample was randomly drawn from a list of fathers who had a child being seen at the clinic and who themselves had gone through the clinic's diagnostic procedure. Each neurotic *S* had been, within the preceding 4 mo., judged to have a neurotic disturbance and to be at 3 or 4 on the 5-point scale of "intensity of emotional disturbance" discussed above. None of the disturbed Ss had received formal psychotherapy. The Ss in both of the normal groups were randomly chosen from the class lists of schools in a middle-class area. The normal adult Ss were selected from among parents of school children to make them similar to the disturbed adult sample in that all of the adult Ss had at least one young child. The mean age of the normal adult group was 37.2; of the disturbed adults, 39.7. This difference was not significant, according to a *t* test ($.20 < p < .30$). After the selection of normal Ss, school guidance counselors were asked to eliminate from the list potential Ss whose families had a history of emotional disturbance which had come to the attention of the school. These judgments resulted in the elimination of 2 adolescent Ss and 1 adult *S*.

In the process of soliciting the cooperation of Ss there were two refusals of cooperation by disturbed boys and no refusals by normal boys or their parents. Among the disturbed adults, one man refused to cooperate, and among the normal adult Ss seven declined. All of this latter group stated that their reasons were based upon conflicts and pressures of time, rather than on reluctance to participate.

RESULTS

Analysis of variance of the *Q*-sort consistency scores (Table 1) revealed both row and column effects and a significant interaction ($p < .05$), indicating that the instrument used had differentially discriminated the groups. Following Lindquist (1953), *t* tests were employed to test the simple effects of interest. Table 2 presents the mean *Q*-sort consistency scores and *t*-test comparisons of the two adolescent groups and the two adult

TABLE 1
ANALYSIS OF VARIANCE OF Q-SORT
CONSISTENCY SCORES

Source of variation	df	MS	F
Age (A)	1	.166	5.57*
Emotional health (B)	1	.134	4.49**
A × B	1	.134	4.49**
Within	56	.0298	
Total	59		

* $p < .025$.

** $p < .05$.

groups. The mean consistency score of the neurotic adolescent group (.192) was significantly lower than the normal adolescents' mean of .318 ($p < .03$). The mean consistency score of the neurotic adult group (.307) was not significantly different from the normal adults' score (.392), demonstrating support for the hypotheses that neurotic adolescents describe themselves less consistently over situations and that this inconsistency is specifically a characteristic of neurotic disturbance in adolescence, rather than of neurosis in general.

A further comparison was made between the two normal groups. The normal adolescent group's mean consistency score of .318 was not significantly different from the normal adult group's mean of .392.

DISCUSSION

The results support the position that neurotic adolescents tend to describe themselves less consistently over situations than do adolescents or neurotics in general. The finding of such an interaction suggests that generalizations from previous work relating neurosis to self-concept flux are unwarranted, at least when describing neurotic adult males.

TABLE 2
t-TEST COMPARISONS OF Q-SORT CONSISTENCY SCORES

	Mean r^2	t	df
Neurotic adolescents	.192	2.00	28*
Normal adolescents	.318		
Neurotic adults	.307		
Normal adults	.392	1.25	28

* $p < .03$.

In addition, these findings have some relevance to the issue of identity crisis as a part of normal adolescence. Consistent with Engle's finding but not with Erikson's theory, normal adolescents in the present study did not tend to describe themselves less consistently than did normal adults. The results reported apply only to the first of Erikson's three criteria, however, and it is quite possible that a more inclusive approach to identity would have resulted in different findings. (In previous research, however, it has been a nearly universal practice to define identity as the perception of oneself as the same person over time and interpersonal situations.) Perhaps the present study is most relevant to the identity issue in that it points up the necessity for further adolescent-adult comparisons in future studies.

Future research should also examine the developmental aspects of the crystalization of a consistent self-description. The present findings suggest that normal children go through such a crystalization process before adolescence and that the presence of neurotic problems delays this process until after adolescence.

REFERENCES

- ALLPORT, G. W. *Becoming: Basic considerations for a psychology of personality*. New Haven: Yale University Press, 1955.
- BLOCK, J. Ego identity, role variability, and adjustment. *Journal of Consulting Psychology*, 1961, 25, 392-397.
- BRONSON, G. W. Identity diffusion in late adolescents. *Journal of Abnormal and Social Psychology*, 1959, 59, 414-417.
- BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *Journal of Abnormal and Social Psychology*, 1952, 47, 597-606.
- BUTLER, J. M. Factorial studies of client-centered psychotherapy. *Counseling Center Discussion Papers*, 1956, 7(9).
- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & R. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954.
- CARTWRIGHT, R. D. The effects of psychotherapy on self-consistency: A replication and extension. *Journal of Consulting Psychology*, 1961, 25, 376-382.
- ENGLE, M. The stability of the self-concept in adolescence. *Journal of Abnormal and Social Psychology*, 1959, 58, 211-215.

- ERIKSON, E. The problem of ego identity. *The Journal of the American Psychoanalytic Association*, 1956, 4, 56-121.
- GOUGH, H. G., & HEILBRUN, A. B. *The Adjective Check List Manual*. Palo Alto: Consulting Psychologists Press, 1965.
- HEILBRUN, A. B. Parental model attributes, nurturant reinforcement, and consistency of behavior in adolescents. *Child Development*, 1964, 35, 151-167.
- HESS, R. D., & GOLDBLATT, I. The status of adolescents in American society: A problem in social identity. *Child Development*, 1957, 28, 459-468.
- HICKMAN, N. W. The role of self-related-concept discrepancies in personal adjustment. *Dissertation Abstract*, 1959, 19, 2656-2658.
- HOOD, J. J. Consistency of self-concept in adolescence. *Dissertation Abstract*, 1962, 22, 2458-2459.
- HOWARD, L. P. Identity conflicts in adolescent girls. *Smith College Studies in Social Work*, 1960, 31, 1-21.
- KELLY, G. A. *The psychology of personal constructs*. Vol. 1. New York: Norton, 1955.
- LECKY, P. *Self-consistency: A theory of personality*. New York: Island Press, 1945.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.
- MUSGROVE, F. Role-conflict in adolescence. *British Journal of Educational Psychology*, 1964, 34, 34-42.
- PHILLIPSSON, L. K. Personality rigidity as a distinguishing factor between normal and neurotic behavior. *Dissertation Abstracts*, 1955, 15, 878.
- ROGERS, C. R. *Client-centered therapy*. Boston: Houghton-Mifflin, 1951.
- ROSENBERG, M. *Society and the adolescent self-image*. Princeton, N. J.: Princeton University Press, 1965.
- TOLOR, A. Rigidity of self-concept as a mechanism in the maintenance of personality equilibrium, and as an expression of this equilibrium. *Dissertation Abstracts*, 1955, 15, 1121-1122.
- WARNER, W. L. *Social class in America*. New York: Harper, 1960.
- WOLDSTAD, A. M. Identity conflicts in disturbed adolescent girls. *Smith College Studies in Social Work*, 1961, 32, 20-37.
- WYLIE, R. C. *The self concept*. Lincoln: University of Nebraska Press, 1961.

(Received September 14, 1967)

RELIGIOUS AFFILIATION AND PSYCHOPATHOLOGY IN A MEXICAN-AMERICAN POPULATION¹

ARNOLD MEADOW AND LOUISE BRONSON²

University of Arizona

Protestant and Catholic Mexican-American Ss of similar levels of acculturation, education, and socioeconomic background were studied with a variety of instruments. Evaluations of psychopathology, reported here, were derived from the L-R sections of the Cornell Medical Index and behavioral observations. The lower rate of pathological responses by Protestant Ss was attributed to the social support offered by the small, intimate congregations with their strong, paternal leadership and the Protestant doctrines of asceticism and individual responsibility which contribute to impulse control.

Despite the popularity of concepts such as "the Protestant Ethic" and the Protestant "need-achievement" culture, studies attempting to contrast the effects of Protestant or Catholic affiliation in personality structure have been in large part unsuccessful. In fact, one recent work, *Religion in American Culture*, (Schroeder & Obenhaus, 1964) indicates that not only are there no observable differences in personality organization but that followers of these faiths when interviewed give vague, almost identical, answers to questions involving the basic tenets of their faiths. The possibility, however, remains that real differences previously observed may have been blunted in the melting pot of American culture. Might these differences appear if "old-fashioned" Catholics could be contrasted with fundamentalist, sect-oriented Protestants?

METHOD

A small group of recently converted, but relatively unacculturated, Mexican-American Protestants and their Catholic counterparts, provided the authors with Ss meeting the above specifications. These Ss were studied with various instruments: a structured interview, a medical questionnaire, projective devices, value schedules, and behavior indexes. This paper reports the results obtained from the medical questionnaire, the Cornell Medical Index (CMI).

¹This research was supported by United States Public Health Service Grant 5-R11-544-2, Arnold Meadow, director.

²Requests for reprints should be sent to Louise Bronson, Student Counseling Bureau, University of Arizona, Tucson, Arizona 85721.

The Cornell Medical Index (CMI)

The CMI was developed as a screening device for both medical and psychiatric problems, and has been used for this purpose in hospitals and clinics. Moreover, the CMI has been used in recent studies in epidemiology: in an area of rapid culture change in North Carolina (Cassel & Tryroler, 1961), in the rural and urban districts of Zulu (Scotch & Geiger, 1964), among the Eskimo (Chance, 1965), in Indian boarding schools (Nelson, Dawes, Hipple, Jetmalani, & Tadlock, 1964), and among inductees of various ethnic origins (Croog, 1961). The psychiatric section here used reports seven areas of symptomatology. Its questions, which are direct, brief, and non-offensive in nature, use simple "yes" and "no" answers, and Spanish and English language versions have been published.

Subjects

The investigation was conducted in a southwestern city consisting of approximately a quarter of a million inhabitants, 17% of whom are Mexican-American. Fifty-four Protestant Ss were drawn from 11 churches, representing the principal Mexican-American Protestant denominations in the city. These were matched with a Catholic group on the variable of age, country of nativity, education, and socioeconomic status. The members of both groups were located through the religious leaders of their church. All were male and 18 yr. or older.

RESULTS

The principle finding of the study was that the Catholics produced a greater number of pathological responses than the non-Catholic group. The mean M-R CMI score for the Mexican-American Catholics was 9.4; for the Protestants 4.6. A two-tailed *t* test shows this difference to be significant at the .001 level.

Figure 1 shows the contrast between Catholic and non-Catholic groups on the

symptom syndromes tapped by the CMI. The t tests show these differences to be significant at the following levels: habits, .001; inadequacy, ns ; depression, ns ; anxiety, .001; sensitivity, .01; anger, .01; and tension, .01.

The markedly high scores in the areas of "inadequacy" and "sensitivity" should be noted. Typical questions in Section M (inadequacy) are, "Do you get nervous and shakey when approached by a superior?" and "Do you wish you always had someone at your side to advise you?" The high scores on this section obtained by the Mexican-American Ss substantiate clinical and field observations. In general, this appears to be a group with strong feelings of inadequacy produced in part by race prejudice and in part by the difficulty experienced in meeting the demands of a technologically oriented world. Section P, (sensitivity) has such questions as, "Are your feelings easily hurt?" and "Does criticism always upset you?" These questions possibly tap a reaction more frequent in a culture employing shame rather than guilt as a mechanism of social control.

The low scores in the category of "depression" are striking. This finding is supported by clinical evidence. Depression, of the type clinically diagnosed in the general population, is seldom seen in the Mexican-American population. Depressed states do occur in Mexican-American patients, but their lack of severity and short duration has led to the coining of the term "Mexican depression" (Meadow & Stoker, 1965).

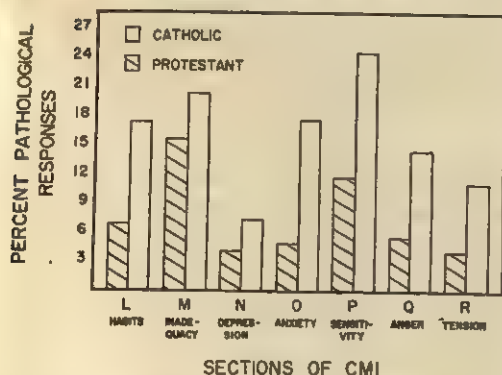


FIG. 1. Catholic and Protestant responses to L-R sections of the CMI.

Although there is a trend for older Ss to give more pathological responses, this tendency fails to reach statistical significance for either Protestant or Catholic respondents.

Examination of the mean scores of Ss of various educational levels reveals a U-shaped relationship (see Figure 2). In separate sign tests it is found that Ss with 0-5 yr. of education have significantly more pathology than those with 6-11 yr. of education (significant at the .01 level), and those Ss with 12 yr. or more of education have more symptomatology than those with 6-11 yr. of education (significant at the .05 level). This relationship may still be observed when age is held constant. It is possible that individuals with limited education simply do not obtain skills requisite for minimal adjustment in an urban American society and consequently develop more psychopathology. Second, it would seem that the most highly educated group is subjected to greater acculturation stress. Representing marginal man caught between the presses of Mexican values stressed by his parents and the new American values learned in school, this group also develops greater psychopathology. Those with average education who have secured sufficient skills for adequate functioning in the American society, but who have not yet encountered to its fullest extent the clash of contrasting cultures, appear to develop the most healthy psychological development.

In drawing conclusions such as those above, the possibility of overgeneralization must be kept in mind. It may *not* be assumed that the differences observed here are also true of Anglo-American Catholics and Protestants. Nor, may these conclusions be applied—without further exploration—to acculturated Mexican-Americans. What happens to these groups as they climb the socioeconomic ladder is an area inviting investigation.

In addition to overgeneralization, the question of response bias must also be considered. Despite the close cultural similarities of the two groups being compared, it is possible that one group might be more motivated to give responses of a certain type. For example, the low scores of the Protestants could be a result of a general tendency to give negative responses or a need to appear in a positive

light. For the following reasons, however, it is felt that the major portion of the differences obtained reflect true differences in symptomatology.

First, if a bias does exist it might be expected to be displayed consistently throughout the CMI responses. There is no significant difference between the responses of Catholics and Protestants on Subsections M (inadequacy) and N (depression) while all other subsections show differences at the .01 or .001 level of significance. Even if one were to assume that the gap between Catholic and Protestant scores on these two subsections represents a base rate of bias, and lowered the differences on the other subscales by this amount, large differences, not accountable to bias, would still remain (see Figure 1).

Second, a few items which ask about family background can be used as a sort of lie scale. For example, to such questions as, "Did anyone in your family ever have a nervous breakdown?" and "Does worrying run in your family?" Catholics and converts to Protestantism should, if no response bias exists, give similar answers. And this is the case. Again, converted Protestants give slightly fewer pathologically oriented responses, but the difference between these responses and those of the Catholics is not significant, even at the .25 level, as reported by a sign test.

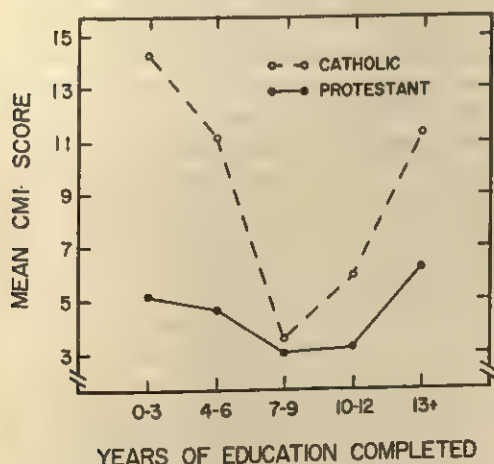


FIG. 2. Catholic and Protestant CMI (M-R) scores and years of education completed.

DISCUSSION

The major finding—lower pathology reported by Protestants—deserves further exploration. What factors have served to thus change this group of converts? Two major influences are proposed as having effected this trend. The first is that the small social groups typical of the Protestant Mexican-American churches provide a greater amount of social support than the larger, more amorphous, Catholic congregations. The second is that the content of the Protestant doctrine itself produces a less pathological character structure than Catholic religious content—at least with respect to the specific problems of adjustment with which the Mexican-American is confronted in an urban American society.

The social support factor may best be evaluated against the background of the Sonoran folk culture from which the Mexican-Americans of this study emigrated. Central to this culture were the supportive institutions of the extended family, the parallel fictional family the *compadrazgo* system, and the economic, political, and emotional support provided by the patron and/or the priest. When the Mexican moves to the United States these traditional institutions are weakened, and no substitutes for these institutional lacunae are readily available to most of the people.

But the Mexican-American Protestant church appears to fill many of these gaps in the weakened traditional culture. Its congregation provides a close knit social system; and, in contrast to the Catholic congregation, it is small, and the members are in continuous social contact. These relationships are further modified in the sect-oriented churches by the shared sense of sin and guilt and the frequent, emotionally fraught, public confessions. Interviews with the 11 ministers and the parishioners of the sampled Protestant churches reveal that the leader of the congregations, himself a Mexican-American, intimately knows the problems of each of his members, and is the patron-priest in a very real sense. In the community chosen for this study the Catholic priests have large parishes and, with one exception, no knowledge of Spanish. They cannot provide the individual

with the multiple social supports which members of a low-status acculturating group so greatly need.

Furthermore, there is the doctrine of the religion itself, the second of the major influences mentioned above. It must be remembered that the Protestants studied here are members of old-fashioned, sect-type churches. Asceticism and individual responsibility are greatly emphasized in these groups and it is these aspects of their doctrine that would appear to be most effective in curbing the types of personality disorganization observed in the Mexican-American community. Basically these problems involve difficulties in control of impulsivity. Crimes of violence, drinking, fighting, impulsive sexual behavior, sudden desertions of family, and the leaving of schools and jobs are among the most frequently observed "problem" behaviors of male Mexican-Americans. In the city under study, the diagnosis of sociopathic personality was the most frequently reported clinical syndrome for the Mexican-American clinic population.

The above hypothesis, linking Protestantism and impulse control, was strongly supported by an interesting behavior observation made during the course of data collection. When the data were tallied on missed appointments, it was found that the Catholic group had failed to meet 57 appointments, while the corresponding Protestant group missed but 6 appointments. A chi-square test shows that this difference is significant beyond the .01 level. Conversations with respondent's wives and assessment of the embarrassed excuses of the Ss themselves led the research team to conclude that the major

variable involved in the appointment cancellations was an impulsive flight from a situation which was perceived as mildly stressful. In some cases, however, the S had quite simply accepted an invitation to a ball game or the corner bar, assuming that the interviewer would understand. The difference between Protestants and Catholics in this respect is not only statistically significant but striking in magnitude. The Protestant Ss, in response to CMI items (particularly in Section Q, anger), indicate that they have increased impulse control; the quantitative data on missed appointments presents objective behavioral data which support their statements.

REFERENCES

- CASSEL, J., & TRYOLER, H. A. Epidemiological studies of culture change. *Archives of Environmental Health*, 1961, 3, 31-39.
- CHANCE, N. A. Acculturation, self-identification, and personality adjustment. *American Anthropologist*, 1965, 67, 378-393.
- CROOG, S. H. Ethnic origins, education level, and responses to a health questionnaire. *Human Organization*, 1961, 20, 65-69.
- MEADOW, A., & STOKER, D. The symptomatic behavior of Mexican-American and Anglo-American hospitalized patients. *Archives of General Psychiatry*, 1965, 12, 267-277.
- NELSON, L. G., DAWES, J. W., HIPPLE, J. L., JETMALANI, N. G., & TADLOCK, L. D. Screening for emotionally disturbed students in an Indian boarding school. Experience with the Cornell Medical Index Health Questionnaire. *American Journal of Psychiatry*, 1964, 2, 1155-1159.
- SCHROEDER, W. W., & OBENHAUS, V. *Religion in American culture*. New York: Free Press of Glencoe, 1964.
- SCOTCH, N. A., & GEIGER, H. J. An index of symptom and disease in Zulu culture. *Human Organization*, 1964, 22, 304-331.

(Received September 15, 1967)

SELF-VERBALIZATION AND EMOTIONAL AROUSAL

DAVID C. RIMM¹ AND STUART B. LITVAK

Arizona State University

The present experiment investigated certain assumptions derived from Ellis' theory relating implicit verbalization to emotional arousal. Emotional responses (galvanic skin response and respiration changes) of *Ss* instructed to silently read sequences of affectively loaded sentences were compared with those of *Ss* who read neutral sentences. The relationship between relevance to *S* of the content of the sentences and level of emotional response was also investigated. In addition, the effect of sentence type (observation, inference, or evaluative conclusion) on level of emotional response was investigated. Results were that *Ss* showed a significantly greater response to affectively loaded sentences than to neutral sentences on two out of three response measures. Neither relevance nor sentence type were found to have any significant effect. These results were interpreted as offering only partial support for the theory.

Behavior theorists agree that neurotic behavioral reactions are intimately related to the individual's learning history; however, there is disagreement as to the mechanism which allows past experience to interact with the current stimulus situation resulting in neurotic behavior. Typically some sort of mediating response is postulated. In the case of phobic reactions, certain theorists (e.g., Mowrer, 1947; Turner & Solomon, 1962) have stressed the role of fear as a mediating response. Mowrer's two-factor theory assumes that by classical conditioning fear becomes attached to a particular stimulus and that avoiding the conditioned stimulus occasions fear reduction which reinforces the avoidance response. While the present authors are in essential agreement with the proponents of two-factor theory, it is not difficult to conceive of instances of avoidance learning which the theory cannot deal with adequately. For example, snake phobias are relatively common in our culture (Lang & Lazovik, 1963) and yet it is rare that the snake phobic can point to a direct aversive encounter (e.g., having been bitten by a snake). Once the fear response has become attached to snake stimulus, it is reasonable to suppose that fear reduction maintains the instrumental avoidance response. But the critical question of how fear became attached to the phobic stimulus in the first place remains unanswered.

A reasonable possibility is that verbal stimuli act as mediators. One can envision a child experiencing physical punishment paired with verbal (auditory) stimuli such as "bad" or "dangerous" such that these cues come to elicit a fear response. Staats, Staats, and Crawford (1962) have shown how neutral words may come to elicit an emotional response as a result of having been paired with aversive stimulation. If at a later time these same verbal cues are paired with visual "snake" cues, one would expect that by higher order conditioning the snake would come to elicit fear. The ease with which evaluative words may function as unconditioned stimuli in a higher order conditioning paradigm has been demonstrated by Staats and Staats (1958).

One theory of neurotic behavior which has relied almost exclusively on verbal mediational constructs is that of Ellis (1962, 1964). Ellis' main theoretical emphasis is on the way in which neurotic behaviors are maintained. It is assumed that such behaviors are maintained by self-verbalizations which serve to elicit negative effect (e.g., fear) and guide instrumental behavior. Presumably emotional responses become attached to objects by higher order conditioning, although this is not discussed in detail. Individuals acquire patterns of self-verbalization by imitating overt verbal behaviors of significant others. Neurotic or self-defeating behaviors result when external stimuli give rise to self-verbalizations which are irrational in nature. By

¹ Requests for reprints should be sent to David C. Rimm, Department of Psychology, Arizona State University, Tempe, Arizona.

irrational Ellis means verbalizations which are either based on arbitrary or unscientific premises or do not logically follow from their premises. Sequences of self-verbalizations may be divided into three categories, in order of occurrence. The first is a simple observation of an environmental happening, for example, "There is a dog in my path." The second is an inferential statement with regard to the observation, "That dog is likely to bite me." The third is an evaluative conclusion, "That would be horrible." Ellis himself classifies implicit verbalizations into two categories, observation and inference *and* evaluation. The authors chose to divide a sequence of sentences into three categories for the purpose of making a more precise analysis.

Irrational thinking (where thinking is equated with self-verbalization) is seen in the inference (dogs don't usually bite people) and in the conclusion (being bitten by a dog, while somewhat painful, is not horrible). Ellis assumes that the anxiety response associated with the preceding sequence originates not with the observation, but with the inference, and the evaluative conclusion. In other words, if one were merely to observe a dog, and not meditate on the possibility of being bitten, and come to no evaluative conclusion, he would experience no marked increase in anxiety.

Using a technique (rational-emotive psychotherapy) which purports to modify self-verbalizations on the part of the patient, Ellis (1962, p. 38) claims an impressively high improvement rate (90% for patients treated for 10 or more sessions). Although Ellis' theory has become increasingly popular over the past several years, and on intuitive grounds is most appealing, it has yet to undergo a rigorous experimental test. The present experiment, while not a comprehensive test of the theory, did examine certain of the more basic assumptions. The most basic assumption is that self-verbalizations can elicit emotional responses. To test this, experimental Ss were instructed to read to themselves triads of sentences, culminating in negative affective conclusions. While Ss engaged in this activity, galvanic skin response and respiration rate and depth were

continuously monitored. Control Ss were required to read to themselves triads of sentences which were affectively neutral and which did not culminate in evaluative conclusions. The theory predicts a greater emotional response on the part of the experimental Ss. The second assumption tested was that there would be marked differences in emotional responding among the three components of a self-verbalized sequence. A test of this involved comparing the mean level of emotional response for each of the three sentences presented to the experimental Ss on a given trial. The theory predicts an appreciably greater response to the third sentence (the evaluative conclusion) and to the second (inference) than to the first (observation).

An additional question, with which the theory does not explicitly deal relates to the relevance of the content of implicit sentences. By relevance is meant the extent to which a S indicates concern over a particular class of experiences. For a S who gives himself a high rating on a fear-of-dogs scale, the dog biting triad would be relevant. To test the role of relevance, prior to the experiment proper, Ss were required to rate themselves on a number of possible problem areas. On the basis of these prior ratings Ss in the experimental group were presented with an equal number of high-relevance and low-relevance triads. On intuitive grounds one would expect a higher level of emotional arousal to high-relevance triads than to low-relevance triads.

METHOD

Subjects

The 27 Ss, both male and female were introductory psychology students at Arizona State University, whose participation fulfilled a course requirement. Nine Ss were assigned to each of two experimental orders and 9 to a control condition.

Apparatus and Materials

Stimulus items were presented via a Kodak Carousel 800 slide projector, projected onto the wall approximately 6 ft. in from of S. The S's respiration rate and depth and galvanic skin response (GSR) were continuously monitored using a Keeler Polygraph, model 6303.

Procedure

Preliminary to performing the experiment, students in an introductory psychology class were asked to complete a rating sheet, containing items possibly related to areas of personal concern. Students rated themselves on a total of 35 scales, with a 0 rating indicating no concern at all, and a rating of 7 indicating that a particular item represented an area of extreme concern or worry. The 35 scales related to possible problem areas including grades, professional achievement, personal appearance, social behavior, sexual behavior or relationships with parents, finances, world situations, and illness and death. When the 95 rating sheets were returned each of the 35 items was ranked according to scale value, and the top and bottom 10 items were selected for use in the experiment proper. None of the Ss used in obtaining the preliminary ratings participated in the experiment.

Upon entering the experimental room, *E* instructed *S* to fill out a rating sheet containing the 20 items (the 10 rated highest and 10 rated lowest in the preliminary study). Seven-point rating scales were again employed. After completing this, the GSR electrodes were placed on *S*'s palm and the pneumograph attachment wrapped around *S*'s chest. The *S* was instructed to get as comfortable as possible. To facilitate relaxation, low intensity semiclassical background music was presented to *S*, via tape recorder. This preparatory period of approximately 10 min. was instituted in order to allow the response measures to stabilize. Five minutes into the preparatory period *S* was asked if he felt comfortable; this uniformly elicited a GSR deflection, which *E* immediately adjusted to a magnitude of 5 mm. for all Ss.

The two items with the highest ratings and two items with the lowest ratings determined which triads of stimuli were to be used with each *S*. For example, suppose an *S* gave the highest rating to concern over grades, the second highest rating to fear of personal injury, the lowest rating to not being noticed by others, and the second lowest rating to a concern with making mistakes. This *S* would then be exposed to the following four triads of sentences:

My grades may not be good enough this semester . . . I might fail out of school . . . that would be awful.

I might get injured or crippled . . . this could make me ugly and incapacitated . . . what an awful thought.

People don't notice me enough . . . maybe there is something unattractive about me . . . how terrible.

I'm always making mistakes . . . I'll never do anything right . . . what a worthless nothing that makes me.

For half of the experimental Ss (nine) the order of presentation was as follows: triad associated with the highest rating presented first, lowest triad presented next, second highest triad presented next, and second lowest triad presented last. For the re-

maining half of the experimental Ss, order of presentation was lowest triad first, highest second, second lowest third, and second highest triad last. Two different orders of presentation were used in order to take into account any possible response habituation effects.

The nine control Ss were also presented with four triads of sentences but the triads consisted of syllogisms of an impersonal nonaffective nature, such as: "Inventors are imaginative . . . Edison was an inventor . . . therefore he was imaginative."

Prior to presentation of the first triad, *S* was told that he would be presented with sets of three sentences, with each member of a given set pertaining to the same general idea. To illustrate, *S* was presented with a neutral syllogism similar to the type presented above. Each *S* was then instructed to read each sentence to himself as it was flashed on the wall in front of him; he was told to concentrate on that sentence and nothing else. The first sentence of the first triad was then presented. Each sentence was presented for 5 sec. with a 10-sec. interval between successive members of a given triad. Successive triads were separated by a 15-sec. interval. This procedure was continued until *S* had been presented with all four sentence triads.

RESULTS

Preliminary Ratings

Of the 35 items rated, the top 10 items, for example, associated with the *most* concern, received a mean rating of 2.92. The item with the highest mean rating (3.96) was "your grades won't be good enough." The lowest of the top 10 items, "You will be embarrassed in public," received an average rating of 2.28. The bottom 10 items received a mean rating of 1.30. The highest of the bottom 10, "You are insincere," had an average rating of 1.60, and the lowest "Your sexual behavior is immoral," was rated at .83.

Habituation

In order to ascertain the general effect on responsivity, of repeated exposure to stimulus items for each response measure, the sum of the response to each of the first six sentences (i.e., first two triads) was compared with the sum of the response to each of the second six sentences (second two triads). For GSR, defined as the maximum deflection in millimeters during the 10-sec. period following the presentation of a given sentence, for the first experimental order (high relevance-low relevance-high relevance-low relevance)

the mean of the difference scores was 1.96 ($t = .27$, ns). For the second experimental order (low-high-low-high), the mean difference score was .78 ($t = .31$, ns).

Similar analyses were performed on the respiration measures. For rate, the response measure was the absolute value of the difference between mean respiration time established over a 10-sec. base-line period immediately prior to the presentation of the first sentence and mean respiration time over the 10-sec. period following presentation of a given sentence. For the first experimental order, the mean of the difference scores (expressed in millimeters rather than seconds) was 1.37 ($t = 1.21$, ns) and for the second order, .33 ($t = .69$, ns). For depth, the response measure was the absolute value of the difference in millimeters between mean depth during base line and mean depth over the 10-sec. period following any given sentence. For the first order, the mean of the difference scores was .52 ($t = .67$, ns) and for the second 1.26 ($t = .91$, ns). For purposes of comparing experimental and control condition, it was necessary to take the absolute (rather than algebraic) value of respiration changes. While Ss tended to be consistent with respect to the direction of change, some showed decreases (in rate or depth) and others, increases. Algebraic values would tend to cancel one another, negating any real differences between experimental and control conditions. While the experimental Ss gave no evidence of habituation on any measure, significant GSR habituation was observed for the controls. The mean of the difference scores was 4.52 ($t = 3.12$, $p < .02$). For respiration rate and depth the means of the difference scores were, respectively, .19 ($t = .58$, ns) and .48 ($t = .91$, ns).

Relevance

To determine the relationship between rated relevance of the content of a particular triad and physiological response to that triad, high- and low-relevance triads were compared. This was accomplished by summing each of S's responses to each sentence associated with both high-relevance triads, and comparing this with the corresponding sum

based upon both low-relevance triads. Since neither experimental group showed habituation effects, (and the groups would not be expected to differ in any other way) they were combined in order to increase the number of degrees of freedom.

For GSR, the mean of the differences was 3.35 ($t = 1.62$, ns). For respiration rate, again using the absolute value of the baseline postsentence difference, the mean of the difference was .30 ($t = .69$, ns). For respiration depth the mean of the differences was .78 ($t = 1.22$, ns). Thus for all response measures, degree of relevance appears to have had little relationship to magnitude of response to a particular triad.

Response to Different Sentences of a Triad

To determine whether experimental Ss responded differentially to the different sentences of triads, analyses of variance were performed for each response measure. Each of S's responses to the first sentence (observation) of each triad was averaged over all four triads, and this was compared with the corresponding average for the second sentence (inference) and the third sentence (evaluative conclusion).

For GSR, the means, based upon all 18 experimental Ss were for objective, inference, and evaluative conclusion sentences, respectively, 10.03, 6.78, and 5.92. While the greater magnitude of response to objective sentence is counter to what the theory predicts, the effect is not significant ($F = 2.77$, ns). For respiration rate, the comparable means were 1.79, 1.99, and 1.92 ($F = .44$, ns). For respiration depth, the means were 2.04, 2.49, and 2.25 ($F = 1.09$, ns). Thus, for all response measures, there is no evidence for differential responding to the different types of sentences. In particular, the hypothesis that there is a greater emotional response to the inference and evaluative conclusion is not supported.

For the controls, for GSR the means for the major premise, minor premise, and conclusion were, respectively, 5.17, 3.56, and 3.94 ($F = 1.96$, ns). For respiration rate the means were .81, .92, and 1.16 ($F = 1.68$, ns) and for depth, 1.17, .89, and .83 ($F = 2.93$,

ns). While this analysis with respect to GSR is somewhat obscured by the significant habituation effect previously noted, in general one can conclude that for the controls as well, responding was not significantly related to sentence type.

Experimentals versus Controls

The final hypothesis to be examined relates to whether the experimental Ss, having been exposed to sentences "affective" in content, showed a greater response than Ss exposed to "neutral" sentences.

To test this a mean score (for each response measure) was calculated, based upon each experimental S's response to each of the 12 sentences presented to him. The combining of data across conditions was permissible since neither habituation, relevance, nor sentence type was found to have any significant effect. Similarly, mean response scores (for the respiration measures) for the control group were calculated, and experimentals and controls were compared. For respiration rate the means for experimentals and controls were, respectively, 1.42 and .69. This difference is significant ($t = 2.20$, $p < .05$). For respiration depth the respective means were 1.69 and .72, representing a highly significant difference ($t = 4.19$, $p < .001$). Such a comparison for GSR is complicated by the habituation observed in the control group. Taking this into account, mean GSRs for the experimental Ss (based upon all 12 sentences) were compared separately with control means based upon the first two triads (first 6 sentences presented to S) and the second two triads (second 6 sentences presented to S). For the first comparison, respective means were 5.68 and 5.55 ($t = .08$, ns). For the second comparison, the difference is in the predicted direction but does not attain significance, with means of 5.68 and 3.13 for experimental and controls, respectively ($t = 1.61$, ns).

Thus for the respiration measures, experimental and controls Ss could be clearly differentiated. For GSR, experimental Ss also showed somewhat greater reactivity though differences were not significant.

Reliability of Measurement

For both respiration rate and depth, the response measure was the absolute value of the difference between base-line response and response following the presentation of a given sentence. The use of absolute values assumed that Ss responded consistently with respect to the direction of this difference. To test this, each of S's responses to the first six sentences were combined algebraically. The S's responses to the second six sentences were treated in the same manner. For respiration rate, the number of experimental Ss showing agreement in sign (i.e., plus to plus or minus to minus) going from the first six sentences to the second six sentences was 14. The probability of obtaining by chance 14 matches in sign, out of a possible 18 is .011. For respiration rate, again 14 out of 18 Ss showed an agreement in sign. For the controls, for rate, 8 out of 9 Ss showed agreement in sign ($p = .017$), and for depth, 7 out of 9 showed agreement ($p = .070$). Thus, with respect to an increase or decrease in response magnitude relative to base line, Ss tended to behave quite consistently.

The analyses of variance, previously discussed in connection with sentence type, provide an additional check on intra-S reliability. For the experimental Ss, for GSR, respiration rate and respiration depth, respectively, the F s based upon individual differences, were 3.40 ($p < .05$), 4.99 ($p < .001$), and 4.04 ($p < .05$). For the controls the F s ($df = 8/16$) were 5.41 ($p < .05$), 24.54 ($p < .001$), and 21.48 ($p < .001$). Thus, Ss tended to be consistent with respect to magnitude as well as direction of response.

Respiration rate and depth measures would not be expected to be independent, and in fact for the experimental Ss the product-moment correlation coefficient was .60 ($z = 2.67$, $p < .01$). For the controls, the correlation shrank to .13 ($z = .37$, ns). The lack of significant correlation for the controls can be attributed to the very limited range of scores, related of course to low magnitude of response encountered in this group.

DISCUSSION

The purpose of the present experiment was to examine certain of Ellis' basic assumptions relating to the effect of self-verbalization on emotional arousal. In fact, self-verbalization was neither directly manipulated nor independently observed. The relevant stimuli directly under *E*'s control were visually presented sentences, although it was assumed that *S* did follow instructions and read each sentence to himself. In this context self-verbalization is a construct having a role quite analogous to that of the "visual image" in Wolpian desensitization (Wolpe, 1958).

Of the assumptions tested, the one most fundamental to the theory received support; *Ss* tended to show a greater emotional response to sentences of an affective nature than to neutral sentences. Since by definition affective stimuli illicit more emotion than neutral stimuli, the result is as much a validation of methodology as a support for the particular hypothesis. The failure to obtain differential effects associated with relevance and sentence type is somewhat more surprising. With respect to relevance, it is unquestionably true that for a given *S*, certain words or sentences have more power as conditioned stimuli than others. It may be that the content areas chosen were such that the average *S* would be expected to have had some negative conditioning associated with each. For the preliminary ratings, the means for the top 10 and bottom 10 differed by only 1.32 scale units, suggesting only fair differentiation. Nevertheless, taking this finding at its face value, it does suggest that inducing an individual to engage in negative self-referential statements produces an emotional response independent of whether he verbalizes the belief that the statement applies to him or does not.

The finding that responses to objective, inferential, and evaluative sentences did not differ significantly is directly counter to what the theory predicts. It might be argued that since *Ss* were free to engage in self-verbalization over the 10-sec. intersentence interval they might well have come to their own evaluative conclusions during this period.

Thus, the response to an objective sentence would have tended to be as great as it was when an inferential or evaluative sentence was presented. Were such the case one would have expected the change in responding to begin late in the intersentence interval. In fact, in most instances the change occurred during or immediately after the presentation of the sentence, suggesting that the response was to the sentence itself. However, as an additional control one might have *Ss* engage in some sort of competing activity, such as counting backwards, (Peterson & Peterson, 1959) over this interval.

It is possible that for young children, highly evaluative sentences containing words such as "awful" or "terrible" function as more powerful conditioned stimuli than sentences of a more objective nature. For the highly verbal college student, however, by a process of higher order conditioning and additional direct conditioning, objective statements could well become just as effect inducing.

In conclusion, the present experiments lend support to Ellis' general contention that self-verbalization has a direct effect on emotional arousal but does not support the specific hypothesis relating to type of sentence. Certain critical questions, which the authors have not dealt with, are worthy of mention. First, what is the relationship between self-verbalization and instrumental behavior? It is reasonable to suppose that once a verbal mediator comes to elicit negative effect or drive, this will motivate instrumental responding. The second question has to do with the maintenance of sequences of implicit verbalizations. One might predict that when such a sequence elicited negative effect, this would function as punishment suppressing the particular sequence. Suppose, however, that the negative effect did drive an instrumental response, culminating in drive reduction. Drive reduction might be expected to reinforce not only the instrumental behavior but also, the verbal mediational response preceding it. This would account for the maintenance of the implicit verbal sequence. Also, by virtue of its own stimulus properties the verbal mediator may become a critical cue in a chain controlling the in-

strumental response (viz., Osgood, 1953). This is within the realm of speculation of course, and the nature of the relationship between implicit verbalization and instrumental behavior is in need of experimental investigation.

REFERENCES

- ELLIS, A. *Reason and emotion in psychotherapy*. New York: Stuart, 1962.
- ELLIS, A. Rational Psychotherapy. In H. J. Eysenck (Ed.), *Experiments in behavior therapy*. New York: Macmillan, 1964.
- LANG, P. J., & LAZOVIK, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology*, 1963, **66**, 519-525.
- MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harvard Educational Review*, 1947, **17**, 102-148.
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford University Press, 1953.
- PETERSON, L. R., & PETERSON, M. J. Short term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193-198.
- STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, 1958, **57**, 37-40.
- STAATS, A. W., STAATS, C. K., & CRAWFORD, M. L. First-order conditioning of meaning and the paralleled conditioning of a GSR. *Journal of General Psychology*, 1962, **67**, 159-167.
- TURNER, L. H., & SOLOMON, R. L. Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn. *Psychological Monographs*, 1962, **76** (40, Whole No. 559).
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford, Calif.: Stanford University Press, 1958.

(Received September 18, 1967)

OPERANT RESPONDING FOR VISUAL STIMULI DURING SENSORY DEPRIVATION:

EFFECT OF MEANINGFULNESS¹

A. MICHAEL ROSSI,² PETER E. NATHAN, ROBERT H. HARRISON

AND PHILIP SOLOMON

Harvard Medical School

Slide series differing in amount of meaningfulness were used as reinforcers for conjugately programmed operant responding during 2-hr. sessions of sensory deprivation and control conditions. Ss were 36 student nurses between the ages of 17 and 22. Every S had the opportunity to respond for the slide series under both sets of conditions. Rates of responding under both conditions were higher for the more meaningful slide series, but rates of responding for the least meaningful reinforcers were higher under sensory deprivation conditions than under control conditions. The results are interpreted as indicating a need for nonmeaningful stimulation as well as for meaningful stimulation during sensory deprivation.

A perusal of the literature on sensory deprivation (SD) research suggests that the major effects of SD are well documented, but the processes underlying these effects still are matters of conjecture. Theories invoked in formulating these conjectures range from neurophysiological to psychoanalytical, but almost all of them assume that various dimensions of normally present exteroceptive stimuli (e.g., variety, complexity, intensity, etc.) are needed by human organisms to function effectively, and that the absence of these dimensions produces a "stimulus hunger" which, if sufficiently intense, gives rise to deleterious effects. This general assumption is supported by diverse studies on stimulus-seeking behavior during SD employing animal Ss (Fox, 1962; Kish, 1955; Moon & Lodahl, 1956; Premack, Collier, & Roberts, 1957; Sackett, Keith-Lee, & Treat, 1963; Wendt, Lindsley, & Adey, 1963) and human Ss (Bexton, 1953; Goldstein, 1965; Jones, 1961, 1964; Jones, Wilkinson, & Braden, 1961; Myers, Murphy, Smith, & Goffard,

1966; Rossi & Solomon, 1964a, 1964b, 1964c, 1964d; Vernon & McGill, 1960; Zuckerman & Haber, 1965).

An unresolved issue is whether the stimulus hunger is for meaningful stimuli only or for nonmeaningful stimuli as well (i.e., for stimulation, *per se*). The purpose of the present study was to provide experimental data bearing on this issue by examining the extent to which meaningfulness differentially affects the reinforcement value of visual stimuli during SD and the extent to which the reinforcement value of meaningfulness differs under control and SD conditions.

METHOD

Subjects

The Ss were 36 paid, volunteer, student nurses in training at the Boston City Hospital. They were recruited by a posted notice soliciting Ss "for a psychological study . . . rate of pay, \$1.50 per hour." Their average age was 18 with a range of 17-22.

Procedure

The SD conditions were created in a 7 × 7 × 8 ft. darkened, sound-attenuated, air-conditioned cubicle used in previous SD research (Rossi & Solomon, 1964a). White noise was conveyed into the room via a wall loudspeaker at a level of loudness that excluded ambient sounds.

The S sat in a reclining lounge chair facing a Plexiglas covered 2 × 2 ft. wall portal whose center was on a horizontal plane with S's eyes. On the other side of the portal was a television monitor

¹This research was accomplished with the support of the Office of Naval Research, ONR Contract Number 1866(S2); ONR Contract Authority NR 142-115.

²Requests for reprints should be sent to A. Michael Rossi, Psychiatry Service, Boston City Hospital, South Department—BB, 745 Massachusetts Avenue, Boston, Massachusetts 02118.

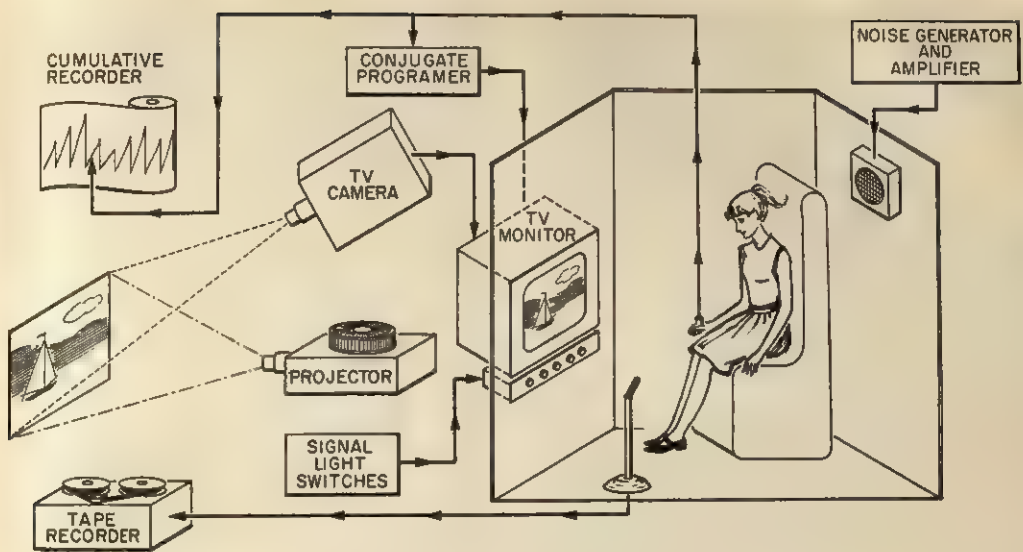


FIG. 1. Schematic diagram of apparatus.

enclosed in a light-proof box and below the portal was a row of 5 colored, 12-v. bulbs. During SD, everything in the room was invisible to *S* except when visual reinforcers were made available. The *S* held a hand switch which controlled (via a conjugately programmed potentiometer) the brightness of the television screen when visual reinforcers were available. During these periods the television screen remained dark if *S* failed to respond, obtained maximum brightness if *S* responded at about two responses (hand-switch closures) per second, and reached intermediate brightness levels with intermediate response rates. The *S*'s responses were recorded on a cumulative recorder. A schematic diagram of the apparatus is presented in Figure 1.

Visual reinforcers were frames of six commercial film strips which portray the home life of typical families in various regions of the world.³ The film strips were composed of sequential photographs containing brief, printed subtitles. Forty frames from each film strip were made into slides to constitute a slide series. This allowed each slide series to be projected in three different sequences to represent the following three degrees of meaningfulness: (a) most meaning was assumed when the slides were projected in their normal sequence; (b) less meaning was assumed when the slides were pro-

jected in a sequence determined by a table of random numbers so that scenes and subtitles did not follow logical orders; and (c) least meaning was assumed when the slides were projected in a random order, upside down, and backwards so that the scenes were barely discernible and the subtitles were illegible.

The six slide series were separated into two groups of three (triads). Every *S* had the opportunity to view both triads; one during a SD session and the other during a control session. In projecting each triad, one slide series was presented in its most meaningful sequence, one in its less meaningful sequence, and one in its least meaningful sequence. Every slide series was projected equally often in each of the three sequences and the design also counterbalanced orders of sessions, orders of meaningfulness within a session, and combinations of triads and sessions.

Following a brief screening interview, *S* was handed a copy of written instructions which described the experimental procedures and emphasized to *S* that *E*s "have no preconceived notion of what you will or should do under these circumstances, so please follow your natural inclinations."

After reading the instructions, *S* was required to operate the hand switch to view a practice slide series of 39 slides in order to become familiar with the equipment and procedure. The practice series contained an equal number of slides in the most, less, and least meaningful sequences. Following the practice period, the experimental sessions began. If *S* was scheduled for the SD control order of conditions, she was placed in SD for 2 hr., and the visual reinforcers were made available during the last half hour of the session. After release from SD, *S* left the laboratory for 1 hr., 25 min. On her return, *S*

³ The filmstrips were produced by Louis de Rochemont Associates for United World Films, Inc., 1445 Park Avenue, New York, N. Y. 10029. The filmstrips used in the present study were: "Eskimo Sea Hunters" (Northwestern Alaska), "Desert Nomads" (French Morocco), "Nomads of the Jungle" (Malaya), "Mountain Farmers" (Switzerland), "Farmer-Fisherman" (Norway), and "Farmers of India" (Middle Ganges Valley).

sat in the experimental room with the door open, the white noise off, and normal sounds audible. After 5 min. for *S* to relax and readjust to the room, the visual reinforcers were made available. The *Ss* scheduled for the control SD order of conditions spent identical amounts of time in SD and out of the laboratory but in reverse sequence. Control and SD sessions were always contiguous.

Identical procedures were used when visual reinforcers were made available during control and SD conditions. The onset of a white warning light signaled to *S* that the visual reinforcers were about to be made available, and *S* had been instructed to operate the hand switch three times to signal *E* that *S* saw the signal light and was not asleep. The white warning light remained on for 10 sec., then one of the colored lights came on and remained on for 5 min., 20 sec. While a colored light was on, a slide series was projected onto a screen in front of a television camera connected to *S*'s television monitor (see schematic diagram in Figure 1). Each slide was projected for 8 sec. The interval between the termination of a colored light and the onset of the next white signal light within a session was 6 min., 45 sec. With three slide series and two intervals between slide series, the total time for the presentation of one triad was exactly $\frac{1}{2}$ hr.

RESULTS

The rate of *Ss*' responses for the visual reinforcers provided the primary data of the study. These data were subjected to a complex analysis of variance with 36 identified sources of variance. The results indicated that only the variance attributed to the different levels of meaningfulness had achieved statistical significance ($p < .001$). A perusal of the mean rates of responding presented in Table 1 indicates that this was attributable to the lower responding for the least meaningful slide series than for the other two series; there were no apparent differences in responding for the most meaningful and less meaningful series.

TABLE 1

MEAN RESPONSE RATES DURING SENSORY DEPRIVATION AND CONTROL SESSIONS FOR VISUAL REINFORCERS WITH THREE DEGREES OF MEANINGFULNESS

Session	Meaningfulness in visual reinforcer			
	Most	Less	Least	Total
Sensory deprivation	694	689	473	619
Control	691	672	373	579
Difference	3	17	100	40

Note.— $N = 36$.

The results of the analysis of variance also indicated that the variance attributed to conditions (SD and control) approached, but did not reach, standard levels of statistical significance ($p < .12$). A graph of the mean response rates for visual reinforcers during both conditions, presented in Figure 2, suggests that the variance attributed to conditions was associated almost solely with differences in responding for the least meaningful reinforcers. The application of Cochran's approximate t test confirmed this suggestion with the findings of no differences in responding for either the most or less meaningful reinforcers between SD and control conditions and significantly more responding for the least meaningful reinforcers under SD conditions when compared to control conditions ($p < .01$).

A visual examination of individual cumulative records indicated that more than a third of the *Ss* showed high variability in rates of response during the presentation of the least meaningful reinforcers, but brief bursts of "stimulus testing" responses (a few responses now and then apparently to determine the nature of the visual reinforcer) were more frequent and sustained longer during SD than during control conditions.

DISCUSSION

The results of this study suggest that different degrees of meaningfulness in visual reinforcers produce different rates of operant responding by *Ss* both in and out of SD. The results also demonstrate that with the response costs (number of responses required to light up the television screen) used in the present study, the reinforcement value of more meaningful visual stimuli did not differ between control and SD conditions, but the reinforcement value of visual stimuli with least meaning was slightly greater during SD than during the control conditions.

The increased reinforcement value of the least meaningful stimuli during SD suggests that a need for nonmeaningful stimulation (i.e., stimulation, *per se*) is involved in the stimulus hunger created by SD. It remains for future research to determine whether the

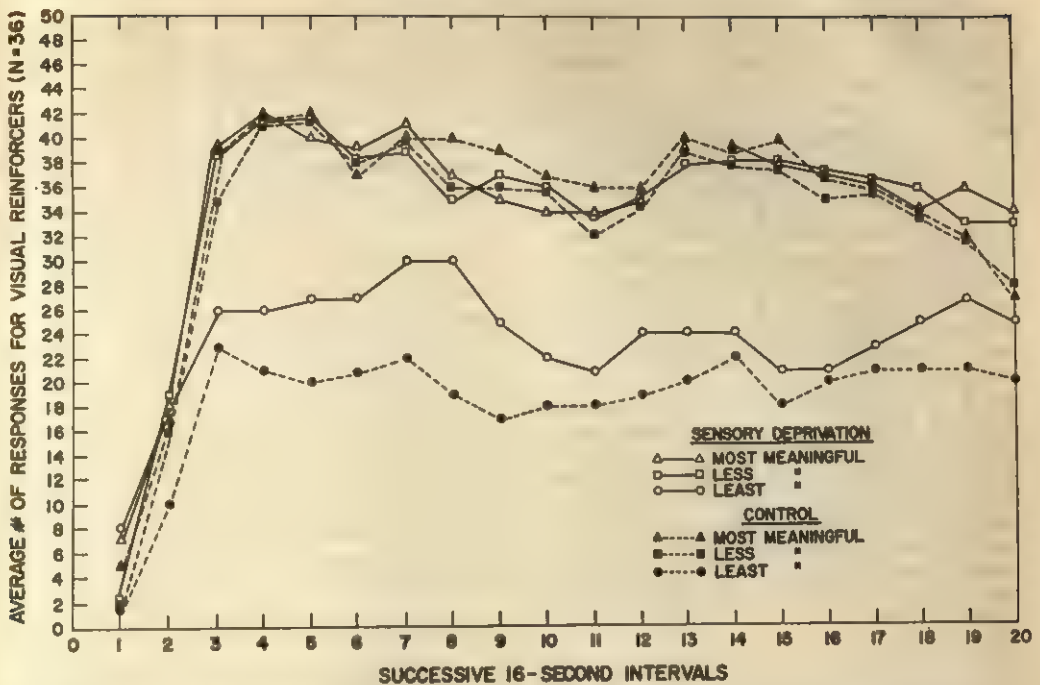


FIG. 2. Button-press responses under control and sensory deprivation conditions for visual stimuli (slide series) with three degrees of meaningfulness. (Each slide in a series was projected for 8 sec.)

differential reinforcement value of meaningfulness can be increased or decreased with higher or lower response costs than were used in the present study and to determine the differential reinforcement value of meaningfulness after more extended periods of SD.

A tangential, but nonetheless important, aspect of this study was the demonstration of a method that provides sensitive, continuous, and automatic recordings of behavior during SD. The versatility of the method makes possible the investigation of the reinforcement value of a wide variety of stimuli ranging from simple stimulus intensity to social interaction via television under controlled and comparable conditions. This method may be able to generate data which will contribute to an objective description of the hypothesized stimulus hunger and the identification of stimulus dimensions which affect its intensity. Hopefully, this will lead to a greater understanding of the processes underlying the deleterious effects observed to occur during SD.

REFERENCES

- BEXTON, W. H. Some effects of perceptual isolation in human subjects. Unpublished doctoral dissertation, McGill University, 1953.
- FOX, S. S. Self-maintained sensory input and sensory deprivation in monkeys: A behavioral and neuropharmacological study. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 438-444.
- GOLDSTEIN, K. M. Stimulus reinforcement during sensory deprivation. *Perceptual and Motor Skills*, 1965, **20**, 757-762.
- JONES, A. Information deprivation and irrelevant drive as determiners of an instrumental response. *Journal of Experimental Psychology*, 1961, **62**, 310-311.
- JONES, A. Drive and incentive variables associated with the statistical properties of sequences of stimuli. *Journal of Experimental Psychology*, 1964, **67**, 423-431.
- JONES, A., WILKINSON, H. J., & BRADEN, I. Information deprivation as a motivational variable. *Journal of Experimental Psychology*, 1961, **62**, 126-137.
- KIBB, G. B. Learning when onset of illumination is used as reinforcing stimulus. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 261-264.

- MOON, L. E., & LODAHL, R. M. The reinforcing effect of changes in illumination on lever pressing in the monkey. *American Journal of Psychology*, 1956, 69, 288-290.
- MYERS, T. I., MURPHY, D. B., SMITH, S., & GOFARD, S. J. Experimental studies of sensory deprivation and social isolation. Technical Report 66-8, June 1966, Human Resources Research Office.
- PREMACK, D., COLLIER, G., & ROBERTS, C. L. Frequency of light contingent bar pressing as a function of the amount of deprivation of light. *American Psychologist*, 1957, 12, 411. (Abstract)
- ROSSI, A. M., FURHMAN, A., & SOLOMON, P. Arousal levels and thought processes during sensory deprivation. *Journal of Abnormal Psychology*, 1967, 72, 166-173.
- ROSSI, A. M., & SOLOMON, P. Button-pressing for a time-off reward during sensory deprivation: I. Compared to activity reward: II. Relations to descriptions of experience. *Perceptual and Motor Skills*, 1964, 18, 211-216. (a)
- ROSSI, A. M., & SOLOMON, P. Button-pressing for a time off reward during sensory deprivation: V. Effects of relatively comfortable and uncomfortable sessions. *Perceptual and Motor Skills*, 1964, 19, 803-807. (b)
- ROSSI, A. M., & SOLOMON, P. Button-pressing for a time-off reward during sensory deprivation: III. Effects of varied time-off rewards. *Perceptual and Motor Skills*, 1964, 18, 794-796. (c)
- ROSSI, A. M., & SOLOMON, P. Button-pressing for a time-off reward during sensory deprivation: IV. Relation to changes in ratings of well-being. *Perceptual and Motor Skills*, 1964, 19, 520-522. (d)
- SACKETT, G. P., KEITH-LEE, P., & TREAT, R. Food versus perceptual complexity as rewards for rats previously subjected to sensory deprivation. *Science*, 1963, 141, 518-520.
- VERNON, J., & MCGILL, T. E. Utilization of visual stimulation during sensory deprivation. *Perceptual and Motor Skills*, 1960, 11, 214.
- WENDT, R. H., LINDSLEY, D. F., & ADEY, W. R. Self-maintained visual stimulation in monkeys after long-term visual deprivation. *Science*, 1963, 139, 336-338.
- ZUCKERMAN, M., & HABER, M. M. Need for stimulation as a source of stress response to perceptual isolation. *Journal of Abnormal Psychology*, 1965, 70, 371-377.

(Received September 18, 1967)

PARENTAL DOMINANCE, CONFLICT, AND DISCIPLINARY COERCIVENESS IN FAMILIES OF FEMALE SCHIZOPHRENICS¹

JOSEPH BECKER² AND HERBERT SIEFKES

University of Washington

Relations between the premorbid adjustment of female schizophrenics and family interactions were investigated. Fifteen good premorbid (GPM), 15 poor premorbid (PPM), and 15 nonpsychiatric control (N-PC) daughters and parents were compared on measures from a structured situational test, parental self-reports, and daughters' retrospective ascriptions of parental attitudes. Predictions of group differences in interparental dominance, and conflict level were weakly supported. N-PC parents attributed more disciplinary coerciveness to themselves ($p < .025$) than parents of schizophrenics. Greatest discrepancy between parents' and daughters' coerciveness ratings occurred in PPM families ($p < .01$), and discrepancy related inversely to quality of premorbid adjustment ($p < .01$). N-PC mothers reported the greatest range of coerciveness ($p < .005$), whereas fathers did not report nor daughters ascribe such differences. PPM daughters ascribed stronger domineering or controlling tendencies to their mothers ($p < .001$) and fathers ($p < .001$) toward themselves than GPMs or N-PCs.

Few studies of schizophrenic family interaction have used female probands, and virtually none include any variant of a good-poor premorbid *S* analysis (Garmezy & Rodnick, 1959). Schematically, the rationale of the present study was that parents of schizophrenics provide their children with poor models for internalizing integrated identities, or for acquiring effective extrafamilial interpersonal instrumental behaviors. Role relations between parents of schizophrenics are characterized by power imbalance, and much conflict. Communication within families of schizophrenics is constricted and fragmented resulting in discrepant perceptions and expectations among family members. More spe-

cifically, this study examined selected aspects of interparental dominance and conflict, and of parent-child coerciveness and dominance as a function of their schizophrenic daughters' premorbid adjustment.

The first hypothesis was that parents of female schizophrenics have an opposite interparental dominance configuration from parents of male schizophrenics (Farina, 1960). That is, families of female good premorbid (GPM) schizophrenics tend to be maternally dominated, those of poor premorbid (PPM) paternally dominated, and those of nonpsychiatric controls (N-PC) relatively equalitarian. The prediction for GPMs derived from reports of female GPMs during conjoint therapy sessions (Becker, 1963, p. 49); the prediction for PPMs from the Fleck, Lidz, and Cornelison (1963) study comparing parental relations of male and female schizophrenics. Fleck et al. stressed the inadequate role models that mothers of inferred PPM females provide; an effect reinforced by their husbands' contemptuous derogation. These mothers were described as low in self-esteem, vague, anxious, depressed, and insensitive, in contrast with their husbands' arrogance, self-investment, and attention seeking. A Bales type analysis of female schizophrenic parental interaction by Cheek (1964)

¹ This investigation was supported by grants from the Psychiatric Training and Research Fund of the Illinois Department of Mental Health, and the University of Washington Graduate School Research Fund Initiative 171 Section. Appreciation is expressed to Ethel Shaeffer, of the Biostatistics Division of the University of Washington Medical School, and to Larry Menninga, of the Computer Center for statistical and programming consultation. Part of the data in this study were included in an unpublished master's thesis submitted to the University of Washington School of Social Work by the second author.

² Requests for reprints should be sent to Joseph Becker, Department of Psychiatry, University of Washington, Seattle, Washington 98105.

gave inferential support to this formulation. The prediction of equalitarian relations between N-PC parents stemmed from Farina's findings.

The second hypothesis was that greatest conflict occurs between parents of PPM schizophrenics and the least between parents of N-PCs. This assumption followed from Farina's (1960) experimental findings, and Fleck et al.'s (1963) description of families of female schizophrenics as characterized by mutual parental undercutting and much overt conflict.

The third hypothesis dealt with parental disciplinary coerciveness or punitiveness. Despite their schismatic relations, many parents of schizophrenics seem preoccupied with apparent compatibility or "pseudomutuality" (Wynne, Ryckoff, Day, & Hirsch, 1958). It was predicted, therefore, (a) that PPM parents report themselves as less coercive disciplinarians than N-PC parents, whereas (b) PPM daughters ascribe more coerciveness to their parents than N-PCs. The greater parental coerciveness ascribed by PPM daughters is construed as reflecting a rupture in familial pseudomutuality (Wynne et al., 1958) which permits more veridical perception. It was also predicted (c) that N-PC parents endorse the broadest range of coercive measures in coping with infractions, and PPM parents the narrowest range. That is, N-PC parents are more likely to discipline by simultaneously restraining, explaining, suggesting, and seeking causes than by harshness, constriction, or distortion as do PPM parents (Lu, 1962; Wynne et al., 1958). Further, the greatest discrepancy between parents' and daughters' perceptions of parental coercive severity was predicted (d) within PPM families, and the least within N-PC families. Investigators of schizophrenic family dynamics have repeatedly stressed the prevalence of faulty communication and of shifting, poorly shared and internalized values (e.g., Ferreira & Winters, 1965).

The fourth hypothesis dealt with each parents' dominance or control in relation to the daughter as retrospectively reported by the latter. It was predicted that PPM daughters ascribe the strongest domineering tendencies toward their parents, and N-PCs the weakest

(Garmezy, Clarke, & Stockner, 1961; Heilbrun, 1960).

METHOD

Subjects

The patient *S* criteria required adequately cooperative, hospitalized females diagnosed as schizophrenic, aged 17-40, Caucasian, neither retarded nor brain damaged, and with their native born natural parents available. Parents were below age 68 and not incapacitated physically. Nonpsychiatric control families met the same criteria as the patients, but had no history of severe emotional disturbances. Controls were obtained through their ministers, and were paid \$10.00. Most families in both groups came from large urban areas (Chicago or Seattle), and were of British, German, or Scandinavian background as judged by surnames.

Schizophrenic patients were rated on the modified Phillips Premorbid Scale (Farina, Garmezy, Zalusky, & Becker, 1962) and the Social Competence Scale (Phillips & Zigler, 1961). The *Ss* below the total patient median of 14 on the Phillips Premorbid Scale, and above the median of .83 on the Social Competence Scale were assigned to the GPM group, while those above the median on the former and below on the latter were placed in the PPM group. Families of three *Ss* whose two scores did not place them in either category, and of three more whose daughters were uncooperative were eliminated. Both scales were used to insure a maximally accurate social maturity dichotomy within a sample of limited size.

Median length of hospitalization for GPMs and PPMs was 2 mo. The study sample included 15 sets of parents with PPM schizophrenic daughters, 15 with GPM daughters, and 15 with N-PC daughters. The group differed significantly on the supplementary variables of age, education, and estimated IQ, but not on socioeconomic class (Hollingshead, 1957).³ Members of the PPM families tended to have less formal education and lower estimated verbal IQs than their GPM and N-PC counterparts.

Procedure

Parents: Private phase. Parents were seen separately, but simultaneously. They filled out a basic data form, the CPI-Fe Scale, Franck Figure Drawings, and WAIS Vocabulary Subtest; parental speech samples were tape-recorded;⁴ and finally, the Jackson Problem Questionnaire (JPQ) (Jackson, 1956) was orally administered following Farina's procedure (1960). The 12 JPQ items require the respondent to indicate how the parent would cope with their child's aggression toward themselves or peers.

³ Hollingshead, A. B. Two factor index of social position. 1957; detailed data are available in an extended report from the senior author.

⁴ Data from these measures are not directly relevant to the concerns of this paper, and therefore will be reported elsewhere.

Parents: Joint phase. The parents were brought together and requested to discuss and reach agreement on each of the JPQ items. Items were read aloud by the *E* who tape-recorded the session.

Daughters. The daughters were seen privately. Sequentially they were administered the WAIS Vocabulary Subtest, Garnezy-Shoben Attitude Scale (G-SAS) (Garnezy et al., 1961), CPI-Fe Scale, speech sampling, Franck Figure Drawings, and the JPQ. On the latter, daughters described what their parents' joint reactions would have been. The G-SAS requires *S* to think back to his early teens and to indicate whether at that time his parents would have agreed or not with 75 statements regarding child-rearing attitudes and practices in the areas of Dominance, Overprotection, and Ignoring. Mother-father items of this scale were presented in counter-balanced order.

Indexes of dominance and conflict. The indexes derived from the JPQ parental interaction were similar to those described in detail by Farina (1960). Scoring of the JPQs was done by an experienced rater with previously well-established interrater reliability.⁵ In addition, the relatively untrained junior author rated four protocols blindly. Interrater agreements were somewhat below those reported by Farina (1960). Spearman rho values for interrater agreement follow the indexes. Dominance indexes were speaks first (1.00), speaks last (1.00), speaks first and last (1.00), yielding maximum (.94), yielding minimum (.88), yielding average (.91), passive acceptance of solution (.95), total time spoken: father minus mother (.93); interruptions: father minus mother (.73), and disagreements and aggressions: father minus mother (.77). Conflict indexes were frequency of simultaneous speech (.92), total duration of simultaneous speech (.95), average duration of simultaneous speech (.93), interruptions-father (.75), interruptions-mother (.74), total interruptions, disagreements and aggressions-father (.71), disagreements and aggressions-mother (.79), total disagreements and aggressions, failure to agree (1.00), verbal activity (.95), and for coerciveness: the maximum (.85), minimum (.55), average (.70), and range (.81) of the private and joint coerciveness scores on the JPQs of the mothers, fathers and ascribed scores of the daughters. Each discrete component of a JPQ response is rated for its coerciveness (Jackson, 1956). Parental dominance scores were those attributed by the daughters to their mothers and fathers on the G-SAS.

RESULTS

Since the groups differed significantly on many of the supplementary variables, corre-

lations between these and the dependent variables involved in testing each hypothesis were examined to determine whether covariance adjustments were needed.⁶

Hypothesis 1. Dominance indexes and supplementary variables were essentially unrelated, therefore each unadjusted index was subjected to a one-way analysis of variance multiple-range test (Dixon, 1965, pp. 572-606). These results are presented in Table 1. The findings afforded quite limited support for the prediction of greater paternal dominance in the PPMs, and greater maternal dominance in the GPMs. On 9 of the 10 interdependent variables, the hypothesized trend of PPM paternal dominance occurred, but the predicted reversal in dominance patterns for PPMs and GPMs attained statistical significance ($p < .05$) for speaks first and last and total time spoken only. The GPM group was the most maternally dominated on all indexes except the yielding ones. Nonpsychiatric controls tended to be most equalitarian with a slight inclination toward maternal dominance.

Hypothesis 2. Conflict scores were adjusted for group differences in socioeconomic class and total time spoken. The prediction that most conflict would occur between the parents of PPMs and least between the N-PC parents received very modest support from the multiple covariate analyses of variance. On 10 of the 11 indexes the PPMs had the greatest amount of conflict, and on 9 of the 11 the N-PCs had the least. But none of the group differences were statistically significant ($p < .05$). Further analysis of these data by Jonckheere's nonparametric test for predicted order of outcome yielded similar results.

Hypothesis 3a. Maximal parental coerciveness was unrelated to the supplementary variables. A 3×3 nested factorial analysis of variance (Dixon, 1965, pp. 503-504) of the groups' (N-PC, GPM, PPM,) JPQ maximum coerciveness scores for mothers' and fathers' private and joint conditions yielded a significant group difference only ($F = 4.70$, $df = 2/84$, $p < .025$). Parents of N-PCs reported the strongest coerciveness (overall \bar{X}

⁵ Eileen Iwakami was the experienced rater who corresponded as necessary with A. Farina over questionable scoring points. Her interrater reliability is discussed in an unpublished paper (Becker & Iwakami, 1967) available from the senior author.

⁶ Tables of intercorrelations are available from the senior author on request.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND MULTIPLE-RANGE ANALYSES OF VARIANCE OF DOMINANCE INDEXES

Index ^a	Group ^b			<i>F</i>	Multiple-range comparisons
	N-PC	GPM	PPM		
Speaks first					
\bar{X}	-.47	-2.53	1.33	1.55	
<i>SD</i>	4.42	6.12	7.20		
Speaks last					
\bar{X}	-.20	-1.73	1.60	1.27	
<i>SD</i>	4.63	5.28	7.02		
Speaks first & last					
\bar{X}	-.67	-4.27	2.93	2.63	PPM > GPM*
<i>SD</i>	7.24	7.00	10.98		
Yield maximum					
\bar{X}	-.34	-.26	.32	.85	
<i>SD</i>	1.58	1.28	1.67		
Yield minimum					
\bar{X}	-.68	-.03	.48	2.82	PPM > N-PC*
<i>SD</i>	.99	1.14	1.76		
Yield average					
\bar{X}	-.55	-.15	.41	2.45	PPM > N-PC*
<i>SD</i>	.87	1.10	1.50		
Passive acceptance of solution					
\bar{X}	.20	-.13	.27	.26	
<i>SD</i>	1.15	1.41	2.15		
Total time speaking					
\bar{X}	20.09	-19.61	162.61	2.77	PPM > GPM*
<i>SD</i>	175.06	243.73	243.05		
Interruptions: father minus mother					
\bar{X}	1.13	-2.27	1.33	1.40	
<i>SD</i>	3.76	9.65	4.97		
Disagreements & aggressions: father minus mother					
\bar{X}	-.07	-1.87	-1.40	.47	
<i>SD</i>	5.13	5.32	5.40		

^a Positive index scores reflect paternal dominance; negative ones maternal dominance.^b *N* = 15 for each group.* *p* < .05.

for N-PCs = 7.23; GPMs = 6.25; PPMs = 6.33) as predicted.

Hypothesis 3b. Daughters' ascribed joint parental coerciveness ratings on the JPQ were unrelated to the supplementary variables. They were therefore subjected to a multiple-range analysis of variance test. Parents of PPMs were rated ($\bar{X} = 7.83$) more coercive ($p < .05$) than those of GPMs ($\bar{X} = 6.57$), but parents of N-PCs were rated ($\bar{X} = 7.03$) intermediately, and not different from the other groups. This hypothesis was not supported.

Hypothesis 3c. Mothers' private self-reported JPQ coerciveness range scores were unrelated to the supplementary variables. A multiple-range analysis of variance of these scores yielded a highly significant *F* of 6.17 ($p < .005$). As predicted, N-PC mothers had

the broadest coercive range ($\bar{X} = 2.81$), PPMs the narrowest ($\bar{X} = 1.75$), and GPMs were intermediate ($\bar{X} = 1.91$). Nonpsychiatric controls differed significantly from GPMs and PPMs ($p < .01$), while the latter two did not differ. Neither fathers' self-reported range scores, nor the parents' joint JPQ range scores, nor the daughters' ascribed JPQ range scores differed significantly. This hypothesis was partly supported.

Hypothesis 3d. Discrepancy scores between parents' joint JPQ and daughters' ascribed JPQ maximum, minimum, and average coerciveness were unrelated to the supplementary variables. As shown in Table 2, the predicted discrepancy between parents' and daughters' perceptions of parental coerciveness was consistently greatest within the PPM families and least within the N-PC families. The

TABLE 2

MULTIPLE-RANGE ANALYSES OF VARIANCE OF DISCREPANCIES BETWEEN PARENT'S JOINT MAXIMUM, MINIMUM, AND AVERAGE JPQ COERCIVENESS AND DAUGHTERS' ASCRIBED JPQ COERCIVENESS

Source	\bar{X}	SD	F^a	Multiple-range comparisons
Maximum coerciveness			3.99*	
N-PC daughters minus parents	-.08	1.52		PPM > GPM & N-PC*
GPM daughters minus parents	.24	2.14		
PPM daughters minus parents	1.63	1.82		
Minimum coerciveness			4.69**	
N-PC daughters minus parents	.11	1.77		PPM > NPC***
GPM daughters minus parents	1.12	1.74		
PPM daughters minus parents	2.08	1.78		
Average coerciveness			5.25***	
N-PC daughters minus parents	-.06	1.41		PPM > GPM* PPM > N-PC***
GPM daughters minus parents	.61	1.72		
PPM daughters minus parents	1.81	1.66		

^a $df = 2/42$.

* $p < .05$.

** $p < .025$.

*** $p < .01$.

schizophrenic groups differed between themselves only on the average coerciveness measure. Discrepancy scores for patient's families correlated .46 ($df = 28$, $p < .01$) with poor-ness of daughters' premorbid adjustment.

Hypothesis 4. Daughters' ascribed G-SAS dominance ratings for fathers and mothers were adjusted for positive correlations with fathers' age ($p < .05$), and subjected to multiple-range analyses of variance. The overall F ($df = 2/42$) for fathers was 8.53 ($p < .001$); fathers were rated most domineering by PPMs ($\bar{X} = 11.41$), least so by N-PCs ($\bar{X} = 5.34$), and intermediately so by GPMs ($\bar{X} = 7.12$). This same pattern occurred in the dominance ratings of mothers (overall $F = 10.81$, $df = 2/42$, $p < .001$; PPM $\bar{X} = 13.07$, N-PC $\bar{X} = 6.14$, GPM $\bar{X} = 7.79$). The GPM and N-PC groups did not differ between themselves for either father or mother, while PPMs differed from both groups for both parents ($p < .01$). There was no difference within any group in strength of dominance ascribed to either parent. The prediction that PPM daughters perceive both parents as having been most controlling was strongly supported.

DISCUSSION

The gains from examining interaction within families of female schizophrenics as a function of the patients' premorbid adjustment were variable. Factors which may have attenuated the findings include the limited

range of premorbid adjustment within the sample, the sensitivity of the premorbid scale to developmental maturity in females, and the questionable validity of the interparental interaction measures (Fontana, 1966). Use of more extreme segments on the continuum of premorbid level and a larger sample size might well bolster the supportive trends on interparental dominance and conflict to statistically reliable levels. However, the dangers of "probability pyramiding" (Neher, 1967) in using multiple indexes from the same interaction source are considerable. Moreover, recently reported low intercorrelations between the dominance indexes from the structured situational test raise serious questions about their assessment of a common construct (Becker & Iwakami, 1967). While considerable construct validation of the premorbid adjustment concept has occurred with male schizophrenics, its sex generality remains less clear. The salient determinants of psychosexual-social developmental maturity in females may differ appreciably from males, and be underrepresented even in the revised Phillips Premorbid Scale.

If Fleck et al.'s (1963) observations on paternal dominance in families of PPM schizophrenics which are weakly buttressed by the data of the present study are valid, it may be speculated that Cheek's (1964) "serrendipitous" finding of greater activity and dominance in her PPM females than controls reflects the tendency of such patients to in-

appropriately overidentify with their fathers.

Comparison of the interparental conflict data of this paper with those of Farina (1960) did not support the Fleck et al. (1963) contention that schismatic (chronically discordant) parental relations are more characteristic of female schizophrenics' families than of males'. Rather, these conflict data indicated the opposite trend. To what extent these conflict findings were a function of the specific task content and structure is indeterminate. The relationship of overt intrafamilial conflict per se to psychopathology does not appear to be a positive linear one in any event (Becker & Iwakami, 1967).

Determination of the basis for the positive relation between severity of pathology and size of discrepancy in parents' reported and children's ascribed coerciveness awaits further investigation. A number of alternative explanations are possible in terms of relations between actual behavioral events, construing of such events and communication and assimilation of such constructions.

Similar ambiguity attends the explanation of the positive relation between severity of pathology and amount of ascribed parental dominance or control. It is unclear, for instance, as to whether PPM parents are actually more controlling, or are perceived in this fashion partially for defensive purposes. Lu (1962), for example, observed that while parents of PPMs are somewhat controlling toward all their children, only the preschizophrenics sought to comply. By exaggerating parental restrictiveness and playing the "good child," the preschizophrenic may defensively avoid anxiety producing awareness of their social inadequacies.

Use of the broadest range of disciplinary methods by mothers of N-PCs should caution against preoccupation with parental consistency as opposed to appropriateness. Subjective inspection of the content of mothers' responses to the JPQ indicated much more stereotypy in the mothers of PPMs. Mothers of N-PCs seemed to take a broader range of antecedent, current, and prospective factors into account when disciplining their children. They would appear to provide their children with models of more flexible and extensive utilization of ego functions.

REFERENCES

- BECKER, J. Good premorbid schizophrenic wives and their husbands. *Family Process*, 1963, 2, 34-51.
- BECKER, J., & IWAKAMI, E. Behavioral ratings and interactions within families of disturbed children. Paper presented at the meeting of Western Psychological Association, San Francisco, May 1967.
- CHEEK, F. E. A serendipitous finding: Sex roles and schizophrenia. *Journal of Abnormal and Social Psychology*, 1964, 69, 392-400.
- DIXON, W. J. (Ed.) *Biomedical computer programs*. Los Angeles: Health Sciences Computing Facility, Department of Preventive Medicine and Public Health, University of California, 1965.
- FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology*, 1960, 61, 31-39.
- FARINA, A., GARMETZ, N., ZALUSKY, M., & BECKER, J. Premorbid behavior and prognosis in female schizophrenic patients. *Journal of Consulting Psychology*, 1962, 26, 56-60.
- FERREIRA, A. J., & WINTERS, W. D. Family interaction and decision making. *Archives of General Psychiatry*, 1965, 13, 214-223.
- FLECK, S., LIDZ, T., & CORNELISON, A. R. Comparison of parent-child relationships of male and female schizophrenic patients. *Archives of General Psychiatry*, 1963, 8, 1-7.
- FONTANA, A. F. Familial etiology of schizophrenia: Is a scientific methodology possible. *Psychological Bulletin*, 1966, 66, 214-227.
- GARMETZ, N., CLARKE, A. R., & STOCKNER, C. Child rearing attitudes of mothers and fathers as reported by schizophrenic and normal patients. *Journal of Abnormal and Social Psychology*, 1961, 63, 176-182.
- GARMETZ, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease*, 1959, 129, 450-466.
- HEILBRUN, A. B. Perception of maternal child rearing attitudes in schizophrenics. *Journal of Consulting Psychology*, 1960, 24, 169-174.
- JACKSON, P. W. Verbal solutions to parent-child problems. *Child Development*, 1956, 27, 339-349.
- LU, Y.-C. Contradictory parental expectations in schizophrenia. *Archives of General Psychiatry*, 1962, 6, 219-234.
- NEHER, A. Probability pyramiding, research error and the need for independent replication. *Psychological Record*, 1967, 17, 257-262.
- PHILLIPS, L., & ZIGLER, E. Social competence: the action-thought parameter and vicariousness in normal and pathological behaviors. *Journal of Abnormal and Social Psychology*, 1961, 63, 137-146.
- WYNNE, L. C., RYCKOFF, I., DAY, J., & HIRSCH, S. Pseudomutuality in the family relations of schizophrenics. *Psychiatry*, 1958, 21, 205-220.

(Received October 10, 1967)

REGRESSION IN SCHIZOPHRENIC THOUGHT DISORDER¹

RAFAEL KLORMAN² AND LOREN J. CHAPMAN

University of Wisconsin

This study investigated the relevance of some of Chapman, Chapman, and Miller's findings concerning schizophrenic verbal behavior to the hypothesis of regression in schizophrenic thought. Chapman et al. found that schizophrenics misinterpret words in context by relying excessively on the strongest normal aspects of meaning, neglecting the weaker aspects of meaning. In the present study, the same task that discriminated schizophrenics from normals in this error tendency was administered to third, fourth, and eighth graders. Younger children differed from older children in the same ways that schizophrenics differed from normal adults. Thus, the hypothesis of regression in schizophrenic thought was confirmed at a descriptive level.

Several clinical writers have stated that schizophrenic thought disorder consists of a regression to infantile levels of thought. This is the theoretical position of such psychopathologists as Arieti (1955), Fenichel (1945), Freud (1950), Kantor and Winder (1959), Levin (1936), and Sullivan (1924). Cognitive theorists who have also endorsed this point of view include Goldman (1962), Piaget (1923), Vigotsky (1934), von Domarus (1964), and Werner (1957). Empirical studies of this hypothesis were reviewed by Chapman, Burstein, Day, and Verdone (1961). The present investigation concerned the relevance to the regression hypothesis of some recent findings of Chapman, Chapman, and Miller (1964) about schizophrenic thought disorder.

In this paper, "regression" is employed merely as a descriptive term referring to similarities observed between the behavior of children and schizophrenics. Thus, regression is not used here as an explanatory principle implying that the patients' thought disorder is due to their having become like children. As Chapman et al. (1961) noted, such a usage would require demonstration of a fairly

blanket similarity between schizophrenics and children.

Following Lewin (1951), regression is used here only to refer to a change to a less mature type of behavior. Regression, then, is not employed synonymously with "retrogression," which Lewin defined as a return to a behavior that was demonstrably present at an earlier stage of a particular individual's development. Empirical demonstration of retrogression would require longitudinal data.

Thus, the criterion used in the present study for inferring regression in schizophrenic thought was that both schizophrenics and children exceed normal adults in the commission of a specific type of error on a given task. The task used in the present study was devised by Chapman et al. (1964) to test their theory concerning schizophrenics' misinterpretation and misuse of words. These authors use the term "meaning responses" to refer to those internal events which represent the denotative referents of a word, and which mediate overt responses to words. The central tenet of this theory is that schizophrenics' misinterpretations of the meanings of words arise in part from mediation of their overt responses to words by the strongest normal meaning responses with a neglect of weaker normal meaning responses. In contrast, normal persons' interpretations of words reflect the use of the weaker as well as the stronger meaning responses. This formulation assumes that meaning responses to a given word as well as their rank ordering on strength are the same for schizophrenics and normals.

¹ This paper is based on a master's thesis (Klorman, 1967) submitted by the first author to the University of Wisconsin and supervised by the second author. This research was supported by a grant (2 T1 MH-5363-18) from the National Institute of Mental Health, United States Public Health Service.

² Requests for reprints should be sent to Rafael Klorman, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

Data supporting this assumption have been provided by Chapman et al. (1964) and Brelje (1967). Normal Ss are presumably able to use either the stronger or the weaker aspects of meaning of a word, whichever may be more appropriate for the task at hand. Schizophrenics, however, often mediate their interpretations of words and their overt responses to words only by stronger meaning responses, regardless of their appropriateness.

This formulation has been successful in predicting schizophrenics' errors in conceptual sorting (Chapman et al., 1964), judgments of synonymity (Chapman & Chapman, 1965), syllogistic reasoning (Gruber, 1965), sorting of objects (Miller, 1967), semantic generalization (Mourer, 1966), and paired-associate learning (Warshawer, 1967).

The study that is most relevant to the present one (Chapman et al., 1964, Experiment I) investigated schizophrenics' misinterpretations of multimeaning words, that is, words with more than one meaning in everyday usage. For example, the word "deck" may mean either "a part of a ship" or "a pack of playing cards." A person receiving a message in which the word deck appears must decide, on the basis of contextual cues, which of its meanings is intended. Chapman et al. (1964) proposed that schizophrenics do not make an adequate response to those contextual cues that indicate that the weaker meaning of a multimeaning word is intended. Instead, the patients inappropriately interpret the word in accordance with its stronger meaning. However, if the stronger meaning of the multimeaning word were intended, schizophrenics' performance would more closely approximate that of normal Ss.

In order to test these hypotheses, the investigators first assessed the relative strength of the meanings of 19 multimeaning words by asking normal judges to indicate and rank order the three most prominent aspects of meaning of these words. An index of the relative strength of the meanings of a word was computed by averaging across judges the ranks for each statement of meaning. (A list of the words and the weights of their stronger and weaker meanings appear in Chapman et al., 1964).

The investigators used these judgments of strength of meaning to construct a test. For each of the 19 multimeaning words, there were two items. One of these required a correct interpretation mediated by a stronger meaning response and offered a misinterpretation mediated by a weaker meaning response. An example follows for the word "bark."

After the noise was over, Janet said, "The bark is bad but it won't hurt you."

This means:

- A. She was talking about the sound made by a dog.
- B. She was talking about the outer covering on a tree.
- C. She was talking about the color of a house.

Alternative A, which is based on the stronger meaning of bark, is correct. Alternative B, which is based on the weaker meaning of bark, is incorrect.

A parallel item required a correct interpretation mediated by a weaker meaning response and offered a misinterpretation mediated by a stronger meaning response.

The gardener said, "Here is an unusual bark."

This means:

- A. He was talking about the sound made by a dog.
- B. He was talking about the outer covering of a tree.
- C. He was talking about the color of a house.

In this item, the correct choice is B, the alternative based on the weaker meaning of bark. Alternative A, which is based on the stronger meaning of bark, is incorrect.

It should be noted that both items contain an irrelevant alternative which has no relationship to the stimulus word bark. In both items above, Alternative C is the irrelevant one; however, in the test as a whole, the irrelevant alternative appeared equally often in all three positions. A certain number of errors might be made for other reasons than misinterpretations mediated by stronger or by weaker meaning responses. For example, an S might check alternatives randomly or on the basis of their position or for other unidentified idiosyncratic reasons. Such unidentified errors presumably would result equally often in the marking of the irrelevant

alternative as in the marking of the other two alternatives. Thus, there should be as many "random errors" hidden within markings of the other two alternatives as within markings of the irrelevant alternative. Therefore, as a correction for such random errors, the number of markings of the irrelevant alternative on items measuring misinterpretations mediated by stronger meaning responses was subtracted from the number of markings of misinterpretations mediated by the stronger meaning responses. The resulting corrected score of the tendency to make misinterpretations mediated by stronger meaning responses is assumed to be relatively uncontaminated by random errors. A comparable correction for random errors was employed for items measuring misinterpretations mediated by weaker meaning responses.

Chapman et al.'s findings confirmed their hypothesis. Although schizophrenics exceeded normals in errors of all types, they deviated more from normals on items in which the correct answer was a weaker response than on items of the reverse type. A difference score between the two types of misinterpretations significantly differentiated the two groups.

In the present study, the hypothesis is that younger children exceed older children on this difference score.

METHOD

Subjects

The Ss were pupils from the third ($N=32$), fourth ($N=33$), and eighth ($N=32$) grades of a parochial school in Madison. Each class was tested as a group during a regular class period and on the same day. Means and standard deviations of chronological age were, third grade: $M=8.72$, $SD=.49$; fourth grade: $M=9.73$, $SD=.41$; and eighth grade: $M=13.59$, $SD=1.37$.

Task

The task used in this study consisted of the 38 items employed by Chapman et al. (1964, Experiment I).⁸ Six Ss circled more than one alternative or failed to circle an alternative on 1 or 2 items. For each such unscorable item, the corresponding paired item was disregarded in the computation of scores for the S involved.

RESULTS

There were no significant between-sex differences within any grade in number of misinterpretations. Therefore, the data for the two sexes were combined at each grade level. Table 1 shows the means and standard deviations of misinterpretations mediated by stronger and by weaker meaning responses and the score of the difference between both types of misinterpretations. All of these scores were corrected for random markings in the manner described above.

As indicated by the Kruskal-Wallis test, the three grades differed significantly in the number of misinterpretations mediated by stronger meaning responses ($H=48.97$, $p<.001$), in the number of misinterpretations mediated by weaker meaning responses ($H=12.46$, $p<.01$), and in the score of the difference between these two types of misinterpretations ($H=35.50$, $p<.001$).

The significance of the differences between pairs of means was tested by Mann-Whitney U tests. Here, as in all the analyses, two-tailed probability values were employed. Third graders exceeded eighth graders in misinterpretations mediated by stronger meaning responses ($z=6.56$, $p<.001$), misinterpretations mediated by weaker meaning responses ($z=4.08$, $p<.001$), and in the score of the difference between these two types of misinterpretations ($z=5.28$, $p<.001$).

Fourth graders exceeded eighth graders in number of misinterpretations mediated by stronger meaning responses ($z=5.03$, $p<.001$). Fourth graders also tended to make more misinterpretations mediated by weaker meaning responses than did eighth graders,

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR
CORRECTED SCORES OF MISINTERPRETATIONS

Error score	Third grade		Fourth grade		Eighth grade	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Stronger	4.34	2.71	2.54	2.60	.22	.42
Weaker	1.25	1.72	.36	1.30	.03	.17
Stronger minus weaker	3.09	2.98	2.18	2.51	.19	.47

Note.—Misinterpretations mediated by stronger and weaker meaning responses.

⁸ A copy of the test is available from either author.

but this difference did not reach statistical significance ($z = 1.35, p < .16$). In the score of the difference between the two types of misinterpretations, fourth graders significantly exceeded eighth graders ($z = 4.76, p < .001$).

Similar effects of age were reflected in the differences between third and fourth graders. Third graders made more misinterpretations mediated by stronger meaning responses ($z = 2.31, p < .03$) and more misinterpretations mediated by weaker meaning responses ($z = 2.19, p < .04$). Third graders tended to exceed fourth graders in the score of the difference between the two types of misinterpretations, but this difference was not significant ($z = 1.32, p < .17$).

The finding that younger children exceeded older children in misinterpretations mediated by weaker meaning responses would appear to contradict the hypothesis that younger children rely on stronger meaning responses at the expense of weaker meaning responses. This finding parallels Chapman et al.'s (1964) results for schizophrenics and normals. These investigators argued that:

the finding that the schizophrenics exceeded the normal Ss in misinterpretations mediated by weaker meaning responses . . . is nevertheless consistent with the theory if the items which contributed to the difference were those on which the judges were less unanimous as to which is the stronger meaning. The logic is that if a meaning response which is weaker by the group judgmental norms is nevertheless the stronger of the two meaning responses to this word for a minority of the judges, it is also stronger for a minority of the schizophrenics [p. 64].

To test this hypothesis, Chapman et al. divided the 19 words into one set of 10 words with relatively high-interjudge agreement (89-100%) as to which of the two statements

of meaning is stronger and another set of 9 words with relatively low-interjudge agreement (50-87%). Schizophrenics significantly exceeded normals in misinterpretations mediated by weaker meaning responses on the 9 items based on words with low-interjudge agreement, but not on the 10 items based on words with high-interjudge agreement.

A comparable analysis was performed in the present study. Table 2 shows the means and standard deviations of misinterpretations mediated by weaker meaning responses on items based on words with high- and low-interjudge agreement.

The three grades did not differ significantly in number of misinterpretations on items based on words with high-interjudge agreement ($H = 1.96, p < .50$). In contrast, the three grades differed significantly in number of misinterpretations on items based on words with low-interjudge agreement ($H = 15.42, p < .001$). On this set of items, third graders exceeded both fourth graders ($z = 2.62, p < .01$) and eighth graders ($z = 4.07, p < .001$) in number of misinterpretations. Fourth graders, however, did not differ from eighth graders ($z = .35, p = .38$). Thus, the difference between third graders and the other two grades in misinterpretations mediated by weaker meaning responses was accounted for by the items based on words with low-interjudge agreement as to which meaning is stronger.

To check for the possibility that the stem sentences in the set of items measuring misinterpretations mediated by stronger meaning responses tended to be more ambiguous than those in the other set of items, ratings of ambiguity of the stem sentences were obtained from 100 introductory psychology students. Items of the following format were presented:

After the noise ceased, Janet said, "The bark is bad, but it won't hurt you."

The word "bark" may mean either

a) "sound made by a dog"

or b) "outer covering on a tree"

In the sentence above, how clear is it that the intended meaning is "sound made by a dog"?

This was followed by a rating scale from 1 ("almost completely clear") to 5 ("not at all clear"). The items for which weak-meaning

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR CORRECTED SCORES OF MISINTERPRETATIONS

Type of word	Third grade		Fourth grade		Eighth grade	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low agreement	1.03	1.42	.21	1.14	.03	.17
High agreement	.22	.94	.15	.56	.00	.00

Note.—Misinterpretations mediated by weaker meaning responses on items based on words with high- and low-interjudge agreement.

responses were correct were rated as slightly more ambiguous ($M = 1.74$) than the items for which strong-meaning responses were correct ($M = 1.65$). This small difference was statistically significant ($p < .001$), but may probably be attributed to the difference in strength of the two meanings of the multi-meaning words, rather than to differences in strength of the contextual cues provided by the sentences in which the words appeared. Nevertheless, subsets of the two types of items were matched on ambiguity ratings ($M = 1.68$) by dropping two items from each set, and the statistical analyses of the study were repeated. The results were substantially unchanged. On the difference score of misinterpretations mediated by stronger meaning responses minus misinterpretations mediated by weaker meaning responses, eighth graders still differed from both third graders ($z = 5.35$, $p < .001$) and from fourth graders ($z = 5.14$, $p < .001$). Similarly, a reanalysis of Chapman et al.'s (1964) data comparing schizophrenics and normal Ss showed that, for the matched subsets of items, this difference score differentiated the groups ($z = 3.55$, $p < .001$). All of the other findings of significant differences in both studies also remained unchanged. These results clearly demonstrated that none of the findings of these two studies were due to a difference in ambiguity of the stem sentences in the items measuring the two types of error.

DISCUSSION

The results of this investigation confirmed the hypothesis that younger children exceed older children in the differential tendency to interpret words in context by relying excessively on the meaning responses which are stronger for normal adults. There is, apparently, a developmental progression in propensity toward this differential error, and the performance of schizophrenics corresponds to an early stage of this developmental trend.

However, one need not conclude from these data that schizophrenics and young children exhibit this differential error tendency for the same reasons. There is, as yet, no clear explanation for schizophrenics' excessive reliance

on stronger normal meaning responses to mediate their overt responses to words. One might speculate that the children who made these errors had not yet acquired the weaker meanings of the words used in the experimental task; however, evidence for this speculation is not presently available.

Nevertheless, the present investigation demonstrated that younger children parallel the tendency of schizophrenics to rely excessively on stronger normal meaning responses in their interpretation of words presented in the context of a sentence. The results of the present study support, at least on a descriptive level, the hypothesis of regression in regard to this aspect of schizophrenic thought disorder.

REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.
- BRELJE, T. The steepness of hierarchies of strength for meaning responses and word associations in schizophrenia. Unpublished doctoral dissertation, Southern Illinois University, 1967.
- CHAPMAN, L. J., BURSTEIN, A. G., DAY, D., & VERDONE, P. Regression and disorders of thought. *Journal of Abnormal and Social Psychology*, 1961, 63, 540-545.
- CHAPMAN, L. J., & CHAPMAN, J. P. Interpretation of words in schizophrenia. *Journal of Personality and Social Psychology*, 1965, 1, 135-146.
- CHAPMAN, L. J., CHAPMAN, J. P., & MILLER, G. A. A theory of verbal behavior in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 1. New York: Academic Press, 1964.
- FENICHEL, O. *The psychoanalytic theory of neuroses*. New York: Norton, 1945.
- FREUD, S. *Collected papers*. Vol. IV. London: Hogarth Press, 1950.
- GOLDMAN, A. E. A comparative-developmental approach to schizophrenia. *Psychological Bulletin*, 1962, 59, 57-69.
- GRUBER, J. The von Domarus principle in the reasoning of schizophrenics. Unpublished doctoral dissertation, Southern Illinois University, 1965.
- KANTOR, R. E., & WINDER, C. L. The process-reactive continuum: A theoretical proposal. *Journal of Nervous and Mental Disease*, 1959, 129, 429-434.
- KLOMAN, R. Regression in schizophrenic thinking. Unpublished master's thesis, University of Wisconsin, 1967.
- LEVIN, M. On the causation of mental symptoms. *Journal of Mental Science*, 1936, 82, 1-27.
- LEWIN, K. *Field theory in social science*. New York: Harper & Row, 1951.

- MILLER, G. A. The reinterpretation and prediction of "concrete" behavior in schizophrenia using the theory of restricted mediation. Unpublished doctoral dissertation, Southern Illinois University, 1967.
- MOURER, S. A. The prediction of patterns of schizophrenic error resulting from semantic generalization. Unpublished doctoral dissertation, Southern Illinois University, 1966.
- PIAGET, J. La pensée symbolique et la pensée de l'enfant. *Archives de psychologie*, Genève, 1923, 18, 273-304.
- SULLIVAN, H. S. Schizophrenia: Its conservative and malignant features. A preliminary communication. *American Journal of Psychiatry*, 1924, 81, 77-91.
- VIGOTSKY, L. S. Thought in schizophrenia. *Archives of Neurology and Psychiatry*, 1934, 31, 1063-1077.
- VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. New York: Norton, 1964.
- WARSHAUER, D. Paired-associate learning in schizophrenics and normals as a function of strongest and weakest shared meaning responses. Unpublished doctoral dissertation, Southern Illinois University, 1967.
- WERNER, H. *Comparative psychology of mental development*. New York: International Universities Press, 1957.

(Received October 2, 1967)

Manuscripts Accepted for Publication in the *Journal of Abnormal Psychology*

- Physiological Effects of Relaxation Training and Hypnotic Suggestion: Gordon L. Paul*: Children's Research Center Building, University of Illinois, Champagne, Illinois 61820.
- Verbal Operant Conditioning of Hospitalized Psychiatric Patients: William Drennen,* William Gallman, and Gene Sausser: Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.
- False Feedback and Longevity of the Conditioned GSR during Extinction: Some Implications for Aversion Therapy: Karl P. Koenig* and David del Castillo: Department of Psychology, University of New Mexico, Albuquerque, New Mexico.
- Anxiety and Overgeneralization: Negative Results: E. Neil Murray*: Department of Psychology, State University of New York, Buffalo, New York 14214.
- Three Faces of Eve Revisited: A Study of Transient Microstrabismus: William S. Condon,* William D. Ogston, and Larry V. Pacoe: Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213.
- Relative Effects of Prolonged Social Isolation and Confinement: Behavioral and EEG Changes: John P. Zubek,* L. Bayer, and Jean Mary Shephard: Department of Psychology, University of Manitoba, Winnipeg, Canada.
- Measuring Impulse Control in Institutionalized Delinquents Using Rorschach Content and Thought Process Scales: Alan H. Roberts* and Robert V. Erikson: Department of Physical Medicine and Rehabilitation, Mayo Memorial Building, University of Minnesota, Minneapolis, Minnesota 55455.
- Learning and Use of Logical Symbols by Deaf and Hearing Subjects: Solis L. Kates*: c/o Mrs. Donna Robbins, Psychology Department, Clarke School for the Deaf, Round Hill Road, Northampton, Massachusetts 01060.
- Effect of Hypnotic Analgesia on Verbal Report and Cardiovascular Responses to Ischemic Pain: John R. Lenox*: Laboratory of Hypnosis Research, 582 Alvarado Road, Stanford, California 94305.
- Combined Test Used for the Diagnosis of Organic Brain Condition: Predictive Validity Based on Radiographic and Electroencephalographic Criteria: J. Kraus*: Callan Park Hospital, Balmain Road, Rozelle, New South Wales, Australia.
- Exaggeration of Self-Aggression Following Alcohol Ingestion in Rhesus Monkeys: A. S. Chamove and H. F. Harlow*: Primate Research Center, University of Wisconsin, 1223 Capital Court, Madison, Wisconsin 53706.
- Effects of a College Student Visitation Program on a Group of Chronic Schizophrenics: Harold M. Buckley, George A. Muench,* and Bernard M. Sjoberg: Department of Psychology, San Jose State College, San Jose, California 95114.
- Recall and Recognition Free Learning in Schizophrenics: Gilead Nachmani and Bertram D. Cohen*: Department of Psychiatry, Rutgers University Medical School, New Brunswick, New Jersey 08903.
- MMPI and Recidivism: James L. Mack*: Norristown State Hospital, Norristown, Pennsylvania 19401.
- Associative and Editing Processes in Schizophrenic Communication: Edward E. Smith*: Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

* Asterisk indicates author for whom address is supplied.

SIMILARITY OF PERSONALITY TRAIT INTERRELATIONSHIPS IN PERSONS WITH AND WITHOUT EPILEPTOGENIC CEREBRAL DYSFUNCTION¹

MELVIN L. SCHWARTZ,² RAYMOND D. DENNERLL, AND YI-GUANG LIN

Michigan Epilepsy Center, Detroit

The problem was posed of whether or not personality tests measure the same traits in brain-injured and non-brain-injured samples. It was hypothesized that, in a sample of persons with epilepsy, personality trait interrelationships would be similar to those interrelationships found in non-brain-injured samples. Using the CPI and EPPS intercorrelation matrices of 199 epileptics, as contrasted with normative sample matrices from both test manuals, the hypothesis was confirmed for the persons with epilepsy. Further confirmation should consider variables such as severity and laterality of pathology, among others. However, insofar as epilepsy is concerned, the implication from Goldstein's work that brain injury *per se* produces a general qualitative dissimilarity in personality trait interrelationships does not appear tenable.

The interrelationship of personality factors in cases of brain injury and also the relationship between various quantifiable aspects of personality in the brain injured has been a seriously neglected area in recent investigations of patients with neurological insults. Klebanoff, Singer, and Wilensky (1954) have pointed out that some of the more traditional psychological tests such as the Rorschach and the MMPI offer little that is new toward the understanding of these relationships. This has been confirmed for personality research in epilepsy by Tizard (1962). An important exception to this conclusion lies in the work of Meier and French (1965) relating the MMPI to laterality dysfunction in psychomotor epilepsy, a disease perfectly suited to this methodology of combined personality-cognitive assessment by virtue of the reportedly high incidence of severe personality disturbances. Certainly an allied but basic issue in the neuropsychology of groups with brain dysfunction, concerns whether or not tests of personality and cognitive functioning are

measuring the same traits and abilities in such a sample as they do in psychopathological or non-brain-impaired (normal) samples.

Goldstein (1940), working with clinical case material, has implied that the interrelationship of thinking processes in the brain injured undergoes qualitative rather than mere quantitative modification. According to this view, behaviors such as cognitive-intellectual abilities or personality traits are supposedly different in kind in brain-damaged as opposed to non-brain-damaged individuals. An investigation by Reitan (1958) undertook to assess the validity of this impression with tests measuring cognitive-perceptual-motor abilities. He found that the Wechsler-Bellevue Form I and Halstead Neuropsychological Battery intercorrelation matrices were essentially similar for a normal and for a brain-injured sample deliberately selected to be heterogeneous for type of brain pathology. Less than 5% of the same pairs of intercorrelations across the two matrices for each test were significantly different. Reitan concluded that "while brain-damaged patients tend to show definite impairment of abilities, their abilities remain essentially of the same kind as those persons without brain damage [p. 345]."

Notwithstanding, Reitan's conclusion has not been tested for its applicability to personality trait interrelationships. The sparse research using factor-analytic methodology com-

¹ This investigation was supported, in part, by Research Grant No. RD-1403-P from the Vocational Rehabilitation Administration, Department of Health, Education, and Welfare, Washington, D. C. Presented in part at the Midwestern Psychological Association meeting, Chicago, May 1967.

² Requests for reprints should be sent to Melvin L. Schwartz who is now at the Department of Psychology, Wayne State University, College of Liberal Arts, Detroit, Michigan 48202.

paring factor loadings in brain-injured and other similar samples (Berger, Bernstein, Klein, Cohen, & Lucas, 1964; Dennerll, Denbroeder, & Sokolov, 1964) is again overwhelmingly devoted to the cognitive-perceptual-motor abilities with the usual neglect of personality variables. The present study, in attempting to extend Reitan's conclusion, investigated the interrelationships among two measures of personality, the California Psychological Inventory (CPI, Gough, 1957) and the Edwards Personal Preference Schedule (EPPS, Edwards, 1953) in a sample of persons with brain dysfunction of epileptogenic etiology and compared these interrelationships with those already established in normative groups on both tests (Edwards, 1953; Gough, 1957). On the basis of clinical observations of the authors made at this epilepsy center it was hypothesized that there would be significant similarities in the personality test intercorrelation matrices of persons with epilepsy as compared to normative samples.

METHOD

Subjects

The Ss were 199 persons with epilepsy (125 males and 74 females) with diagnoses confirmed by neurological examination, seizure history, and EEG. They were a series of consecutive volunteers in an outpatient diagnostic and research center for epilepsy who were participating in a larger study of neurological, social, and psychological factors in the employability of epileptics. The mean age of the group was 26.4 yr. (range 15.0–55.4), mean education was 11.4 yr. of completed schooling (range 7–17), and mean WAIS Full-Scale IQ was 97.7 (range 61–137). The sample was unselected with regard to type of seizure, and contained Ss with grand mal, petit mal, psychomotor, and mixed seizure types. The Ss came from such diverse sources as general practitioners, self-referrals, neurologists, schools, the Department of Vocational Rehabilitation, hospitals, out-patient clinics, employers, and concurrent case files of the Michigan Epilepsy Center.

Procedure

The CPI and EPPS, embedded in a larger battery of psychometric techniques were individually administered to each S. They were scored according to the directions and values provided by the respective manuals. The correlational methodology used by Reitan was also used here to facilitate direct cross-study comparisons. This methodology employed inter- and intratest correlations as the basis for comparison among groups. In the intertest correla-

tions, each of the 18 CPI subscale raw scores was correlated with each of the 15 EPPS subscale raw scores. The consistency score was omitted from consideration in the latter. This provided an intercorrelation matrix containing 270 r 's for Ss with epilepsy. Another matrix for males only with epilepsy was also generated. These matrices were compared with normative data matrices provided by both test manuals. Gough's manual provided an EPPS-CPI intercorrelation matrix for 45 college males of high scholastic aptitude (Gough, 1957; hereafter referred to as the "Gough matrix"). Intratest correlations in the epilepsy sample were established by correlating each CPI subscale with every other subscale on that test. The analogous procedure was carried out for the EPPS. Intrascala correlation matrices for "normal" Ss were also available in the respective test manuals. These normative intratest matrices were then compared with those evolving from the epilepsy sample.

The specific comparisons between matrices were of two types: (a) pair-wise intercorrelation and (b) matrix correlation. In the pair-wise comparisons, every *intertest* subscale correlation in the normative samples was compared with the same correlation in the epilepsy samples. In the matrix comparisons, the CPI-EPPS matrix of the one sample was correlated, using each r as a score, with the same matrix in the other sample, yielding one r which expressed the overall degree of relationship between the two matrices.

RESULTS

The intercorrelation of CPI and EPPS scores yielded an 18×15 matrix of 270 r 's for male epileptics separately, and for all (male and female) epileptics combined.^a There were 96 significant r 's (.05 level used in this and subsequent tests) in the male epilepsy group matrix and 110 significant r 's in the combined-sexes matrix. The CPI-EPPS matrix in Gough's manual yielded only 28 significant r 's, 19 of which were also significant in the male epilepsy matrix. The nearly four times as many significant r 's in both epilepsy matrices relative to Gough's data may lead to the deceiving impression that EPPS and CPI traits are more closely related in samples of epileptic Ss than in a non-brain-injured sample. Of course, for purposes of statistical significance, each correlation was con-

^a A 33×33 matrix combining intra- as well as intertest correlations in the epileptic group could have been generated. However, since the Gough manual did not contain either an EPPS or CPI *intratest* correlation matrix for the 45 males on whom it reported CPI-EPPS *intercorrelations*, a contrast of epileptics and these normal males on intratest r 's was prohibited.

sidered relative to the parameters (mean, variation) of only the sample from which it was derived. Hence, the above values do not reflect any comparison between identical pairs of correlations across the two matrices. Such pair-wise comparisons would be required in order to discover the actual magnitude and thence, the significance of the differences between each pair of correlations in the comparison of any two matrices.

For purposes of pair-wise comparisons, the 270 r 's in each CPI-EPPS matrix were converted to z values using Fisher's r to z transformation (Edwards, 1960). The z value for each r in each matrix was tested for significance against the corresponding z value of the same r in the other matrix. Comparison of the matrices of the normative sample males as compared to epileptic males yielded 16 or 5.9% of pairs with differences significant at or below the .05 level. By chance alone, 5% of the pairs would be expected to differ significantly, suggesting that the differences between corresponding correlations in the two matrices are essentially negligible. Comparison of the Gough matrix with an all female epilepsy matrix yielded 15 or 5.6% significant differences and with the combined male and female epilepsy matrix, 12 or 4.4% significant differences. These latter two difference tests may be somewhat artificial in that they compare female epileptics with normal males. Nevertheless, they do suggest that neither the epilepsy variable nor the sex variable are sufficient to produce major trait interrelationship differences between the normals and the epileptics.

Although the lack of differences between epileptics and normals is certainly the primary finding of this study, there did remain four identical interrelationships in each of the three epilepsy matrices (male, female, and combined sexes) which were significantly different ($p < .05$) from the corresponding Gough matrix r 's. The interrelationship between Autonomy (EPPS) and Intellectual Efficiency (CPI) in Gough's sample was +.26, whereas in the three epilepsy samples, the respective values were -.18, -.25, and -.13. The three remaining significant correlational differences were: Abasement (EPPS)-Dominance (CPI): epilepsy r 's = -.36, -.37,

-.40; Gough: +.01; Nurturance (EPPS)-Dominance (CPI): epilepsy r 's = -.13, -.15, -.14; Gough: +.36; Nurturance (EPPS)-Sociability (CPI): epilepsy r 's = -.15, -.18, -.14; Gough: +.26. Interpretation of so few identical significant differences would be premature if not hazardous without a replication, however.

Finally, the correlation between entire matrices was examined. Using each correlation in each matrix as a score, the Gough matrix correlated .47 with the male epilepsy matrix, .44 with the female epilepsy matrix, and .51 with the combined male and female epilepsy matrix. These intermatrix r 's are not as high as those reported by Reitan for cognitive tests (.79, .78, .85). Personality trait intercorrelations, however, are usually lower than the intercorrelations of cognitive or intellectual abilities. Nevertheless, these three intermatrix r 's are all statistically significant ($p < .001$) and represent a sufficient degree of agreement to state that there is a striking similarity in personality trait interrelationships in the epileptic and normative samples. Further support for this conclusion was also found in the total matrix correlations of .92 and .85 (both $p < .001$) for normal males versus epileptic males and normal females versus epileptic females, respectively, between the CPI 18 \times 18 intracorrelation matrix for each group. Similarly, for the 15 \times 15 EPPS intracorrelation matrices, the same comparisons yielded r 's of .79 and .82 (both $p < .001$).

DISCUSSION

The results demonstrate very striking similarities between a group of Ss with epileptogenic dysfunction and two normative non-brain-damaged groups in the interrelationship of CPI and EPPS personality traits. Although there may very well be equally striking quantitative differences between this sample of Ss with brain dysfunction and the normal samples used here,⁴ the presence of brain dys-

⁴ Either test may differentiate epileptics from non-epileptics on elevation (mean subscale values) or certain specific scale interactions, even though, as these results imply, the internal factorial structures are not at variance for these groups. However, this kind of differentiation, if it exists, is not the focus of this study.

function did not modify the correlational interrelationships between the various traits measured by the two tests. Therefore, Reitan's finding of essentially similar interrelationships between cognitive abilities in a brain-damaged versus a non-brain-damaged sample can be extended to the area of personality traits in cases of brain dysfunction of an epileptogenic variety. The results do not offer any support for those theories of the epileptic personality which posit that there are unitary, consistent, or characteristic personality features of all forms of the disease. However, as Tizard (1962) has concluded in her review of the theories of the epileptic personality, "progress in this field depends on a recognition and study of the complex environmental and pathophysiological factors involved, and on the development of reliable criteria with which to classify epileptics [p. 207]."

Certainly the results of this study are limited in generalization to the epilepsies. Further validation of these findings must await studies manipulating the massive network of variables underlying various neurological insults. Site and type of lesion, severity, and laterality of pathology, for example, must all be varied systematically, with similar results, before a sweeping generalization of the qualitative similarity of personality traits and

abilities of a brain-damaged to a non-brain-damaged population is made.

REFERENCES

- BERGER, L., BERNSTEIN, A., KLEIN, E., COHEN, J., & LUCAS, G. Effects of aging and pathology on the factorial structure of intelligence. *Journal of Consulting Psychology*, 1964, 28, 199-207.
- DENNERLL, R. D., DENBROEDER, J., & SOKOLOV, S. L. WISC and WAIS factors in children and adults with epilepsy. *Journal of Clinical Psychology*, 1964, 20, 236-240.
- EDWARDS, A. L. *Manual: Edwards Personal Preference Schedule*. (Rev. ed.) New York: Psychological Corporation, 1953.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.
- GOLDSTEIN, K. *Human nature*. Cambridge, Mass.: Harvard University Press, 1940.
- GOUGH, H. G. *Manual: California Psychological Inventory*. (Rev. ed.) Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- KLEBANOFF, S. G., SINGER, J. L., & WILENSKY, H. Psychological consequences of brain lesions and ablations. *Psychological Bulletin*, 1954, 51, 1-41.
- MEIER, M. J., & FRENCH, L. A. Some personality correlates of unilateral and bilateral EEG abnormalities in psychomotor epileptics. *Journal of Clinical Psychology*, 1965, 21, 3-9.
- REITAN, R. Qualitative versus quantitative mental changes following brain damage. *Journal of Psychology*, 1958, 46, 339-346.
- TIZARD, B. The personality of epileptics: A discussion of the evidence. *Psychological Bulletin*, 1962, 59, 196-210.

(Received October 4, 1967)

MULTIDIMENSIONAL ANALYSIS OF "HYPNOTIC" BEHAVIOR¹

THEODORE XENOPHON BARBER² AND DAVID SMITH CALVERLEY

Medfield Foundation, Harding, Massachusetts

Ss who had not previously participated in a hypnotism experiment were told simply to close their eyes for 5 min. and place themselves in hypnosis. These Ss were almost as responsive as a group of Ss who were exposed to a formal hypnotic-induction procedure on four dimensions of "hypnotic" behavior: (a) trancelike appearance (e.g., apparent limpness-relaxation); (b) responses to test suggestions to hallucinate, to be amnesic, etc.; (c) reports of unusual experiences such as reported "disappearance" of the body or body parts; and (d) testimony of having been hypnotized. An additional group of Ss (controls) who were told only to close their eyes for 5 min. were generally less responsive than the place-yourself-in-hypnosis group and the hypnotic-induction group on the dimensions of hypnotic behavior.

During the past decade a wide variety of Ss have been exposed to repeated suggestions that they were becoming relaxed, drowsy, sleepy, and were entering a trance state. Typically, these repeated suggestions gave rise to one or more of the following sets of behaviors: (a) Ss appeared passive, limp, very relaxed, as if they were in a trance; (b) they manifested a rather high level of response to standardized test suggestions (e.g., suggestions for limb rigidity, hallucination, and amnesia); (c) they reported that they had unusual experiences; and (d) they testified that they were hypnotized. The authors believe that a general theory is needed that will explain each of these four sets of behaviors that are typically elicited by procedures that are commonly labeled as hypnotic inductions. However, the authors are not as yet convinced that a theory which succeeds in explaining these behaviors will necessarily employ constructs such as *hypnosis*, *hypnotized*, or *hypnotic trance*. Furthermore, it is the authors' opinion that a satisfactory theory will be based on empirical studies which (a)

establish reliable methods for measuring each of the four sets of behaviors, (b) determine the strength of association between the behaviors (R-R laws), and (c) specify the antecedent variables that are functionally related to each of the behaviors (S-R) laws. Before an investigation is presented that was concerned with the latter aims, the presentation will be clarified by describing more precisely the four sets of behaviors that need to be explained. These behaviors, which the authors are labeling as dimensions of "hypnotic" behavior, are as follows:

1. Trancelike appearance. Weitzenhoffer (1957, p. 210; 1963, p. 42) has described four facets of the trancelike appearance that are typically manifested by Ss who have been exposed to a hypnotic-induction procedure, namely, (a) a type of limpness-relaxation, (b) loss of spontaneity, (c) psychomotor retardation, and (d) "trance stare." In the investigation presented below the authors will delineate methods for measuring each of these four components of trancelike appearance and will attempt to relate each component to antecedent variables that are present in the situation.

2. Response to test suggestions. This set of behaviors is elicited by directly stating (suggesting) to S that he will experience certain sensations or events or that he will perform certain behaviors. For instance, it is suggested to S that he cannot bend his arm, he will see and hear objects (that are not present), and he will forget everything that occurred. The set of behaviors that is elicited by suggestions

¹ This research was supported by grants (MH-07003 and MH-11521) from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Mildred O'Brien of Worcester City Hospital, Phyllis Olson of Memorial Hospital of Worcester, and A. Simoens of Rhode Island Medical Center for urging their students to participate in the experiment and for providing experimental rooms at their respective hospitals.

² Requests for reprints should be sent to Theodore Barber, Research Department, Medfield Foundation, Harding, Massachusetts 02042.

of this type is by far the best quantified of the four under consideration. Recent work by Weitzenhoffer and Hilgard (1959, 1962), Shor and Orne (1962), London (1963), and Barber (1965) has yielded a series of standardized scales which satisfactorily measure response to test suggestions for limb rigidity, hallucination, amnesia, analgesia, age regression, posthypnotic behavior, and so on.

3. Reports of unusual experiences. Gill and Brenman (1959, pp. 13-17) have presented a detailed description of the verbal behaviors that are subsumed under this dimension. They state that Ss who have been exposed to a hypnotic-induction procedure typically report (a) alterations in the size of the body or body parts, (b) changes in equilibrium, (c) feeling as if part of the body or the entire body had disappeared, (d) feelings of unreality, (e) changes in experienced temperature, and (f) feeling as if the E's voice is either very near or very far. In the investigation to be described the authors will attempt to quantify and to relate to antecedent variables each of these six types of reports of unusual experiences.

4. Testimony of having been hypnotized. Hilgard and Tart (1966, p. 205) have underscored the importance of this dimension, writing that Ss' reports that they entered a hypnotic state "furnish a somewhat crude but useful answer to the need for evidence that a person is hypnotized. . . ." Methods for measuring this dimension have been presented in several recent studies (Barber & Calverley, 1966; Hilgard & Tart, 1966; Tart & Hilgard, 1966).

The experiment presented in the following section was designed (a) to quantify each of the four dimensions of hypnotic behavior, (b) to ascertain the interrelations among the four dimensions, and (c) to isolate antecedent variables that play a role in determining to what degree S will respond on each dimension. One question at the forefront of discussion is, Are the four sets of hypnotic behaviors intercorrelated? For instance, if a S is highly responsive to test suggestions, does he also manifest a trancelike appearance, report many unusual experiences, and testify that he was deeply hypnotized? Another question is derived from the following consideration:

Hypnotic-induction procedures include several distinguishable antecedent variables. For instance, S (a) sits quietly with eyes closed for a period of time and (b) is told directly or indirectly that he is expected to enter hypnosis. Are the four sets of hypnotic behaviors functionally related to these two distinguishable antecedent variables that are included in hypnotic-induction procedures? An additional question derives from the fact that Ss have different preexperimental expectations concerning how they will perform in a hypnotic experiment. Can Ss' performance on each of the four behavioral dimensions be predicted by assessing their expectations prior to the experiment?

METHOD

Overall Experimental Design

The Ss were assigned to three groups (two experimental groups and one control group). Each S was tested individually by one E (DSC). The E was not acquainted with Ss prior to their entrance into the experimental room.

The Ss assigned to one of the experimental groups were exposed individually to a standardized hypnotic-induction procedure (hypnotic-induction group). The Ss allocated to a second experimental group were told individually to close their eyes for 5 min. and place themselves in hypnosis (place-yourself-in-hypnosis group). The Ss allocated to a third group were individually told only to close their eyes for 5 min. (controls).

The Ss assigned to the hypnotic-induction and place-yourself-in-hypnosis groups were assessed pre-experimentally on expectations (self-predictions) of hypnotic depth. After imposition of the experimental and control treatments, all Ss were rated individually on trancelike appearance as indicated by limpness-relaxation, loss of spontaneity, trance stare, and psychomotor retardation. The hypnotic-induction and place-yourself-in-hypnosis groups were also rated on response to test suggestions, for example, arm immobilization, hallucination, and amnesia. Postexperimentally, all Ss filled out a questionnaire which assessed reports of unusual experience and testimony of having been hypnotized.

Recruitment of Subjects

All Ss were female student nurses. They were "coerced volunteers" in that they were told by their supervisors that they were expected to participate in the experiment in order to become better acquainted with psychological research (cf. Boucher & Hilgard, 1962).

Experimental subjects. The nursing supervisors at two hospitals directed their students, 112 in all, to report individually at scheduled times to a specified

room to participate in a psychological experiment involving hypnosis. Two Ss failed to keep their appointments. The remaining 110 were assigned randomly, in order of their appearance in the experimental room, to one of the two experimental groups (hypnotic-induction or place-yourself-in-hypnosis group) with 55 to each group.

Control subjects. A control group was included in the experiment in order to determine to what extent the four sets of hypnotic behaviors are due to the fact that Ss sit quietly for a period of time with eyes closed. If the control Ss were to serve this purpose, it was imperative that they should *not* have interpersonal contacts with the experimental Ss. If the control and experimental Ss talked with each other, it appears very likely that the controls would learn from the experimental Ss that the investigation involved hypnosis and, consequently, the controls might interpret their instructions (to close their eyes for 5 min.) as instructions to place themselves in hypnosis. To prevent this contamination, the control and experimental groups were recruited from different institutions; that is, the controls were 50 student nurses, who were in training at another hospital, and who were directed by their supervisors to report individually at scheduled times to a specific room to participate in a psychological experiment. All 50 met their appointments.

It is worth noting that, although the control and experimental groups were in training at different hospitals, they came from a homogeneous population. The control and experimental Ss were of the same sex (female), of the same age (17–22), training for the same occupation (nursing), raised in the same geographical locality (eastern Massachusetts and Rhode Island), and of similar ethnic-religious background (predominantly Irish and Italian Catholic).

Expectations of Hypnotic Depth

After each S in one of the two experimental groups was greeted and asked her name and age, she was given a mimeographed preexperimental questionnaire. One item on this questionnaire asked whether S had previously permitted anyone to try to hypnotize her. A second item was worded thus:

I expect (predict) that in this experiment I will be (please fill in percentages so as to sum to 100%):

- (a) deeply hypnotized _____% of the time
- (b) medium hypnotized _____% of the time
- (c) lightly hypnotized _____% of the time
- (d) not hypnotized _____% of the time

Total = 100%

This expectations-of-hypnotic-depth item was scored thus: The percentage figures placed by S in parts (a), (b), (c), and (d) were multiplied by 3, 2, 1, and 0, respectively, and the resultant figures were added and divided by 100. Thus, S's score was 2.0 if she expected that she would be medium hypnotized 100% of the time and her score was 1.0 if she expected to be lightly hypnotized 100% of the time.

Since the authors did not want the control Ss to define the experimental situation as hypnosis, they were not given the preexperimental questionnaire.

Experimental and Control Conditions

Upon completing the preliminaries described above, the Ss were treated as follows:

Hypnotic-induction group. The 55 Ss allocated to this group were exposed individually to Form A of the Stanford Hypnotic Susceptibility Scale (SHSS, Weitzenhoffer & Hilgard, 1959, pp. 13–26). Specifically, each S assigned to the hypnotic-induction group was first exposed to statements intended to establish rapport and to suggestions of postural sway and then was exposed to a hypnotic-induction procedure comprised of repeated suggestions of eye heaviness, eye closure, relaxation, deep sleep, and deep hypnosis.³ Upon completion of the hypnotic-induction procedure, and without a break in the continuity of presentation, each S was assessed on response to the 10 final test suggestions that are included in the SHSS: namely, hand lowering, arm immobilization, finger lock, arm rigidity, hands moving together, verbal inhibition, hallucination of a fly, eye catalepsy, posthypnotic response, and amnesia.⁴

Place-yourself-in-hypnosis group. The 55 Ss assigned to the place-yourself-in-hypnosis group were not exposed to the hypnotic-induction procedure. Each S in this group was first told individually that her participation in the hypnosis experiment could contribute to scientific knowledge, she was next assessed on the postural sway test (from the SHSS), and then was instructed as follows:

"During the next five minutes I want you to close your eyes and place yourself in hypnosis. Let yourself go as deeply into hypnosis as you possibly can and wait for further instructions."⁵

On completion of the 5-min. period, the final 10 test suggestions included in the SHSS were administered.

Control group. As stated above, 50 control Ss were included in the experiment to determine the effects

³Two minutes before the end of the hypnotic-induction procedure each S in this group was told, "I want you to open your eyes, to remain in hypnosis, and to wait for further instructions. Now open your eyes and wait for further instructions." The E observed the S's eye movements for 15 sec. and rated trance stare in accordance with the criteria presented in the next section of this paper. The E then stated, "Close your eyes," and continued the hypnotic-induction procedure for an additional two min.

⁴The E read the SHSS to S with the exception of four paragraphs that were presented from memory. The memorized portion was presented when E was rating Ss on *trancelike appearance* (see next section of this paper).

⁵Trance stare was rated during the fourth minute of the 5-min. period in the same way as for the hypnotic-induction group (see Footnote 3).

due simply to sitting quietly for a period of time with eyes closed. Soon after each *S* in the control group had entered the experimental room, she was individually instructed thus:

"I want you to close your eyes during the next five minutes. Now close your eyes and keep them closed and wait for further instructions."^a

Trancelike Appearance

According to Weitzenhoffer (1957, 1963) and other authorities (Erickson, Hershman, & Secter, 1961, pp. 53-58; Pattie, 1956), hypnotized *Ss* typically manifest a trancelike appearance. As indexes of trancelike appearance these authors list (a) limpness-relaxation, (b) loss of spontaneity, (c) trance stare, and (d) psychomotor retardation. Each of the four indexes was measured in the present experiment. Each *S* was asked to open her eyes for 15 sec. and trance stare was scored during this eyes-open period (see Footnotes 3, 5, and 6). Limpness-relaxation and loss of spontaneity were assessed during the 3-min. period which preceded the eyes-open interval and psychomotor retardation was rated toward the end of the experimental session. Specifically, the following criteria for scoring each of the four indexes of trancelike appearance were derived from Weitzenhoffer's (1957, pp. 210-211; 1963, p. 42) detailed description of the phenomena.

Limpness-relaxation was considered to be present if *S* appeared relaxed or was slumped in the chair, her head was tilted downward toward the chest or backward or sideways over a shoulder, and her hands either hung limply on the sides or rested limply on a support (Weitzenhoffer, 1957, p. 210). The *E* used a stopwatch to record the amount of time *S* met these criteria. Scores of 3, 2, 1, and 0, respectively, were assigned if, during the 3-min. period which preceded the instructions to open the eyes for 15 sec., *S* manifested limpness-relaxation more than 90% of the time, 50% to 90% of the time, less than 50% of the time, and not at all.

To score loss of spontaneity *E* recorded the number of times, during the 3-min. period preceding the instructions to open the eyes for 15 sec., that *S* spoke, opened her eyes, or voluntarily changed position of body or limbs (Weitzenhoffer, 1957, pp. 210-211). A score of 3 was assigned for loss of spontaneity if, during the 3-min. period, *S* did not speak, open her eyes, or change position of body or limbs. If *S* changed position of body or limbs once, she was given a score of 2; if she changed position twice, she was given a score of 1; and if she spoke, opened her eyes, or changed position of body or limbs more than twice, she received a score of 0.

^a To rate trance stare the controls were told, during the fourth minute of the 5-min. period, "I want you to remain like you are and open your eyes. Now open your eyes and wait for further instructions." The *E* observed the *S*'s eye movements for 15 sec. and, after rating trance stare, asked *S* to close her eyes again.

To score trance stare *E* carefully observed *S*'s eye movements during the 15-sec. period when her eyes were open during the experiment. In accordance with the criteria presented by Weitzenhoffer (1963, p. 42), trance stare was considered to be present if *S* did not move her eyes around the room; that is, if she "stared" at or focused upon a point in space or at a particular object. Scores of 3, 2, 1, and 0, respectively, were assigned if trance stare was present more than 90% of the time, 50% to 90% of the time, less than 50% of the time, and not at all.

Immediately before *E* stated that the experiment was over, *Ss* in the control group were told to extend the right arm straight out and then to extend the left arm straight out. The *Ss* in the experimental groups were also told to extend the right and left arms straight out (at the beginning of the hand lowering and the arm rigidity test suggestions of the SHSS). The time in seconds between the end of the instructions to extend each arm and the completion by *S* of movement of the arm to horizontal position was carefully measured with a stopwatch. The score for psychomotor retardation was the average of these two measurements. Thus, slower arm movements (longer time) indicated more psychomotor retardation (Weitzenhoffer, 1963, p. 42).

Reliability of ratings of trancelike appearance. The ratings of trancelike appearance were found to be reliable ($p < .001$) in a separate pilot study with three individuals (who did not participate in the experiment that is described in this paper). When asked to participate in the pilot study, these individuals (one medical technician and two secretaries) were told (a) they would each be exposed 10 times to a hypnotic-induction procedure, (b) they were to act, in each session, as if they became hypnotized to a deep, medium, or light level, and (c) they were to choose, prior to each session, which level of hypnosis they would simulate without informing the investigators as to the level they had selected. During each of the 30 individual sessions, two investigators (the co-authors of this paper) rated *S* on the indexes of trancelike appearance. Satisfactory interobserver agreement was obtained. Using the criteria delineated above, the interrater reliabilities (Pearsonian correlations), averaged over the 30 sessions, were .87, .99, .92, and .91 for limpness-relaxation, loss of spontaneity, trance stare, and psychomotor retardation, respectively.

Reports of Unusual Experiences

Immediately after it was stated that the experiment was over, each of the 160 *Ss* was given a mimeographed postexperimental questionnaire. The first part of this questionnaire included six items that were designed to assess unusual experiences such as changes in the size of the body and feelings of unreality. In constructing these items the authors adhered very closely to Gill and Brenman's (1959, pp. 13-17) description of the phenomena. The items were worded as follows:

1. During the experiment, did you experience any alterations in the size of your body or in the size of parts of your body? For instance, did your whole body or your legs, arms, head, mouth, lips, etc. feel large or swollen or small or shrunken? Yes_____ No_____.

2. During the experiment did you experience any changes in equilibrium such as the following: sensations of giddiness or dizziness; as if you were rocking back and forth; as if the chair was sliding away from you; as if your body had lost its moorings and was floating up near the ceiling; or as if you were sinking endlessly down into space? Yes_____ No_____.

3. At some time during the experiment did you feel as if parts of your body (your arm, foot, leg, etc.) or your entire body had "disappeared," that is, as if you had lost parts of your body or all of your body? Yes_____ No_____.

4. When you were asked to open your eyes during the experiment, did things seem unreal, shadowy, dreamlike, or as if covered by a veil or haze? Yes_____ No_____.

5. At any time during the experiment did you suddenly feel very cold or very hot? Yes_____ No_____.

6. At any time during the experiment did you feel as if the experimenter's voice was very close (as though he were talking into your ear) or as if the experimenter's voice was going further and further away? Yes_____ No_____.

Each of the six items received a score of 1 for a yes answer and a score of 0 for a no answer.

Testimony of Having Been Hypnotized

The second part of the postexperimental questionnaire included the following item: "I believe that during the experiment I was: deeply hypnotized _____% of the time; medium hypnotized _____% of the time; lightly hypnotized _____% of the time; not hypnotized _____% of the time." This testimony-of-having-been-hypnotized item was scored in the same way as the expectations-of-hypnotic-depth item that was included in the preexperimental questionnaire.

The postexperimental questionnaire also included an additional item: "What basis did you use to judge your depth of hypnosis?" Also, Ss in the place-yourself-in-hypnosis group were asked, "How did you go about placing yourself in hypnosis?"

Immediately after completing the postexperimental questionnaire, each S was told that the experiment would be ruined if she discussed it with the other Ss before all had been tested. Each S stated that she understood the importance of this admonition and promised to adhere to it.

Perception of the Situation by the Control Group

A series of precautions were taken to prevent the control group from defining the situation as hyp-

nosis. For instance, the controls did not come in contact with the experimental Ss (to whom the situation had been defined as hypnosis), they were not given the preexperimental questionnaire which assessed expectations of hypnotic depth, and they were not exposed to the rapport instructions, the postural sway test, the hypnotic-induction procedure, and the test suggestions that are included in the SHSS. To determine whether the precautions were successful, Ss in the control group were treated as follows when they were assembled in class the day after the experiment was completed.

The controls were asked to write the answers to several questions. Since it appeared likely that anonymity would produce more truthful responses, they were told not to sign their names. Three of the questions, which were written on the blackboard were:

- (a) Write everything you heard about the experiment before you were tested.
- (b) During the experiment, when you eyes were closed, did you at that time think the experimenter was trying to hypnotize you or wanted you to be hypnotized?
- (c) Did you think that the experiment might have been concerned with hypnosis prior to the time when you saw the question on hypnosis on the questionnaire you received at the end of the experiment?

In reply to the first question, no S mentioned the word hypnosis. One S answered yes to the second and third questions, and the remaining 49 answered no. The authors concluded from these data that the efforts to prevent the control Ss from defining the situation as hypnosis were quite successful. The answers to the above three items strongly indicated that, with one exception, the 50 Ss in the control group did *not* believe, when they were sitting quietly with eyes closed for 5 min., that they were supposed to enter hypnosis or that they were being tested for hypnosis.

Previous Participation in a Hypnotism Experiment

The preexperimental questionnaire, which was given to the 110 Ss in the two experimental groups, asked if S had previously permitted anyone to try to hypnotize her. The 50 control Ss were asked the same question after the experiment was completed. Four of the 110 Ss in the two experimental groups and none in the control group stated they had been previously exposed to hypnosis and that the hypnosis was relatively unsuccessful—they either were not hypnotized at all or were hypnotized only to a very light level. In the present experiment, 2 of the 4 Ss who had prior experience were assigned to the place-yourself-in-hypnosis group and 2 to the hypnotic-induction group. Since the performance of each of the 4 experienced Ss on the dimensions of hypnotic behavior was either slightly below the mean or very close to the mean performance of the inexperienced

Ss in their respective groups, the results for the place-yourself-in-hypnosis group and the hypnotic-induction group will be presented without making special distinctions between the 4 experienced and the 106 inexperienced Ss in these groups.

Analyses of Data

The scores obtained on each dependent variable by the hypnotic-induction, place-yourself-in-hypnosis, and control groups were subjected to one-way analysis of variance. When the variance analysis yielded a significant effect, it was further localized by performing Duncan multiple-range tests upon the group means. Several variables that were measured only for the two experimental groups were tested for significance by *t* tests. Also, Pearson product-moment correlations were computed among each of the major variables. These intercorrelations were calculated separately for each of the treatment groups.

RESULTS AND DISCUSSION

Comparison of Control Group with Experimental Groups

Both the control group and each of the experimental groups were assessed on trance-like appearance, reports of unusual experiences, and testimony of having been hypnotized. Table 1 shows that the control group did not differ significantly from either the hypnotic-induction or the place-yourself-in-hypnosis group on one of the indexes of trancelike appearance (loss of spontaneity)

and on three of the six reports of unusual experiences (alterations in the size of the body or body parts, changes in equilibrium, and changes in experienced temperature). These results indicate the following proposition which might well be taken into consideration in formulating a theory of hypnotic behavior:

Proposition 1. Some of the behaviors that have been traditionally associated with the word hypnosis—for example, loss of spontaneity, and verbal reports of having had some types of unusual experiences (such as alterations in the size of the body, in equilibrium, and in experienced temperature)—may be due simply to the fact that *S* sits quietly for a period of time with eyes closed.

Although Ss in the control group manifested a few hypnotic behaviors, they differed from Ss in the experimental groups on most of the behavioral dimensions. As Table 1 shows, the control group obtained significantly lower scores than either of the experimental groups on three of the indexes of trancelike appearance (limpness-relaxation, trance stare, and psychomotor retardation). Also, the control group obtained lower scores than both experimental groups on three of the six items which assessed reports of unusual experiences

TABLE 1

MEAN SCORES (AND STANDARD DEVIATIONS) ON PREEXPERIMENTAL EXPECTATIONS AND ON FOUR DIMENSIONS OF HYPNOTIC BEHAVIOR

Behavioral dimensions	Experimental groups		Control group
	Hypnotic-induction	Place-yourself-in-hypnosis	
Preexperimental expectations (self-predictions) of hypnotic depth	1.5 _a (.9)	1.5 _a (.9)	
Trancelike appearance			
Limpness-relaxation	2.2 _a (1.0)	1.5 _b (1.0)	.7 _a (.3)
Loss of spontaneity	2.6 _a (.8)	2.3 _a (1.0)	2.6 _a (.8)
Trance stare	1.8 _a (1.2)	1.3 _b (1.2)	.9 _a (.9)
Psychomotor retardation	2.9 _a (1.8)	1.8 _b (.8)	1.3 _b (.4)
Response to 10 test-suggestions	5.8 _a (2.9)	4.8 _b (2.9)	
Reports of unusual experiences			
Alterations in size of body or body parts	.51 _a (.54)	.47 _a (.49)	.36 _a (.47)
Changes in equilibrium	.85 _a (.35)	.76 _a (.42)	.66 _a (.46)
"Disappearance" of body or body parts	.20 _a (.40)	.16 _a (.36)	.04 _b (.20)
Feelings of unreality	.69 _a (.45)	.69 _a (.45)	.26 _b (.43)
Changes in experienced temperature	.36 _a (.47)	.42 _a (.48)	.34 _a (.46)
E's voice seemed very close or far	.71 _a (.44)	.64 _a (.47)	.34 _b (.46)
Testimony of having been hypnotized	1.5 _a (1.0)	1.5 _a (1.0)	.3 _b (.5)

Note.—Standard deviations are in parentheses. Means in the same row containing the same subscript letter do not differ from each other at the .05 level.

TABLE 2

PERCENTAGE OF SUBJECTS TESTIFYING THAT THEY WERE HYPNOTIZED TO A DEEP, MEDIUM, LIGHT, OR LESS THAN LIGHT LEVEL

Self-ratings of hypnotic depth	Hypnotic-induction group	Place-yourself-in-hypnosis group	Control group
Medium deep to deep (scores of 2.5-3.0)	25%	25%	0%
Medium to medium deep (scores of 2.0-2.45)	15%	15%	2%
Light to medium (scores of 1.0-1.95)	31%	25%	12%
Less than light (scores of .10-.95)	22%	23%	20%
Not hypnotized (scores of .0)	7%	11%	66%

("disappearance" of the body or body parts, feelings of unreality, and *E*'s voice seeming very near or very far). (It is worth noting, however, Table 1, that one-fourth of the controls reported that "things seemed unreal" and one-third reported that *E*'s voice seemed either very close or very far.) Also, the great majority of the controls rated themselves as not having been hypnotized whereas the great majority of *S*s in both experimental groups rated themselves as having been hypnotized to some degree (Table 2). In summary, these data suggest the following proposition.

Proposition 2. Although a few behaviors that have been traditionally associated with the word hypnosis can be attributed to the fact that *S* sits quietly for a period of time with eyes closed, the following hypnotic behaviors cannot be clearly attributed to this factor: (a) trancelike appearance as indicated by limpness-relaxation, trance stare, and psychomotor retardation, (b) reports of certain types of unusual experiences such as feelings of unreality, and (c) testimony of having been hypnotized.

How do Inexperienced Subjects "Place Themselves in Hypnosis?"

The reader will recall that 55 *S*s were told to place themselves in hypnosis and that 53 of these had not previously participated in a hypnotic experiment. Postexperimentally, *S*s in this group were asked, "How did you go about placing yourself in hypnosis?" Answers to this question could be classified into four categories. These categories (and the percentage of *S*s giving answers within each category) were as follows: (a) "didn't do anything" or "thought of other things" (13%); (b) "just relaxed," "tried to relax,"

or "tried to relax and to go to sleep" (57%); (c) tried to make my mind a blank" or "tried not to think of anything" (7%); and (d) "tried to relax" and also "tried not to think of anything" (22%). These data suggest two propositions that appear relevant to a theory of hypnotic behavior:

Proposition 3. Present-day *S*s who have not had prior hypnotic experiences have clear-cut preconceptions pertaining to what hypnosis is supposed to involve.

Proposition 4. When asked to place themselves in hypnosis, *S*s who have not had prior hypnotic experience typically try to relax and, in some instances, also try not to think of anything.

Comparison of Place-Yourself-in-Hypnosis Group with Hypnotic-Induction Group

Is a formal hypnotic-induction procedure necessary to elicit a trancelike appearance, high response to test suggestions, reports of unusual experiences, and testimony of having been hypnotized, or can one elicit these behaviors to the same degree by simply instructing *S*s (who have not previously participated in hypnotic experiments) to place themselves in hypnosis? The data summarized in Tables 1 and 2 indicate that the place-yourself-in-hypnosis group did not differ significantly from the hypnotic-induction group in reports of unusual experiences and in testimony of having been hypnotized and that the two groups differed only to a small degree in trancelike appearance and in response to test suggestions.

Reports of unusual experiences and testimony of having been hypnotized. Table 1 shows that *S*s who were told to place themselves in hypnosis did not differ from those

that received the formal hypnotic-induction procedure on any of the six items which assessed unusual experiences. The Ss in both groups reported the same degree of experienced alterations or disappearance of the body or body parts, changes in equilibrium, feelings of unreality, etc. Furthermore, Table 1 shows that, on the average, Ss in the place-yourself-in-hypnosis group and also those in the hypnotic-induction group rated their depth of hypnosis as between light and medium ($M = 1.5$). Table 2, which presents the frequency distribution of the self-ratings of hypnotic depth, shows that 25% of the Ss in each of the experimental groups (and none in the control group) rated themselves as having been deeply or medium deeply hypnotized (self-ratings of 2.5–3.0). In brief, these data suggest the following proposition:

Proposition 5. With inexperienced Ss (who have not previously participated in hypnotic experiments), the statement "Place yourself in hypnosis" is as effective as a formal hypnotic-induction procedure in eliciting reports of unusual experiences and testimony of having been hypnotized.

Trancelike appearance. Table 1 shows that the two experimental groups did not differ significantly on one of the indexes of trancelike appearance—loss of spontaneity. However, Table 1 also shows that the hypnotic-induction group manifested more trancelike appearance than the place-yourself-in-hypnosis group as indicated by the remaining three indexes—limpness-relaxation, trance stare, and psychomotor retardation. Although the differences on the latter three indexes were statistically significant, they were not very large. For instance, the average time required to move the arm to a horizontal position (the measure of psychomotor retardation) was about 3 sec. for the hypnotic-induction group and about 2 sec. for the place-yourself-in-hypnosis group. It appears possible that the small difference on three of the four indexes of trancelike appearance were due to the following: the hypnotic-induction group, but not the place-yourself-in-hypnosis group, received repeated suggestions of relaxation, drowsiness, passivity, and sleep. These suggestions may have implied to Ss that they were to become limp and relaxed, they were

not to move their eyes actively around the room (trance stare), and they were expected to carry out limb and body movements slowly (psychomotor retardation). Further research is needed to ascertain if this hypothesis is more valid than other possible hypotheses in accounting for the differences.

Response to test suggestions. Both experimental groups were assessed on response to the final 10 test suggestions of the SHSS. The Ss exposed to the formal hypnotic-induction procedure typically passed 1 more test suggestion than those told to place themselves in hypnosis (Table 1); on the average, the hypnotic-induction group passed about 6 of the 10 test suggestions ($M = 5.8$) whereas the place-yourself-in-hypnosis group passed about 5 ($M = 4.8$). The 1-point difference in suggestibility between the two groups, which was statistically significant ($p < .05$), could have been produced by anyone of a number of antecedent variables that were included in the hypnotic-induction treatment but not in the place-yourself-in-hypnosis treatment. For instance, the former treatment but not the latter included (a) motivational instructions ("keep up your attention and interest and continue to cooperate. . . . Nothing will be done that will cause you any embarrassment. Most people find this a very interesting experience. . . .") and (b) suggestions that S could now experience unusual effects (e.g., "You are going to experience many things that I will tell you to experience"). Further studies are needed to determine whether it was one or both of these antecedent variables or other variables that produced the 1-point difference in suggestibility.

In summary, the data presented in the preceding two paragraphs suggest the following proposition:

Proposition 6. A formal hypnotic-induction procedure is a little more effective than the statement "Place yourself in hypnosis" in eliciting a trancelike appearance and in facilitating response to test suggestions.

There are two considerations that pertain to the data presented in this section and in the preceding one:

1. Adults in Western culture are well acquainted with the traditional connotations

of the word hypnosis (London, 1961; Orne, 1959). Their familiarity with the traditional lore appears to play an important role in determining their performance in a hypnotism experiment. When present-day Ss who have not previously participated in a hypnotism session are told to place themselves in hypnosis, that is, when they are asked to use what they have previously heard about hypnosis to structure their behavior, they typically perform as follows: they try to relax and they manifest a trancelike appearance, show a rather high level of response to test suggestions, report unusual experiences, and testify that they were hypnotized; that is, they typically behave very much in the manner traditionally expected of "hypnotized" Ss.

2. The statement "Place yourself in hypnosis" appears to be sufficient to produce most of the effects associated with formal hypnotic-induction procedures. This experimental outcome raises several questions with respect to previous treatises which (a) presented elaborate or involved techniques for "inducing hypnosis" (Erickson, 1948; Weitzenhoffer, 1957), or (b) derived elaborate theoretical formulations from complex manipulations involved in formal hypnotic-induction procedures (Gill & Brenman, 1959; Kubie & Margolin, 1944). The present data appear to make it incumbent upon those who emphasize the importance of the manipulations involved in formal hypnotic-induction procedures first to demonstrate rigorously what specific effects are produced by hypnotic-induction procedures that are not produced by the simple statement "Place yourself in hypnosis" and then to delineate empirically which of the many variables in-

cluded in formal hypnotic-induction procedures are relevant and which extraneous in producing each of the effects.

Preexperimental Expectations as Predictors of Hypnotic Performance

With respect to both the hypnotic-induction and the place-yourself-in-hypnosis groups, there were small positive correlations between Ss' preexperimental expectations (self-predictions) of hypnotic depth and the degree to which they responded to test suggestions, reported unusual experiences, and testified post-experimentally that they were hypnotized (Table 3, Row 1). The correlations between preexperimental expectations and trancelike appearance were also in the positive direction but were very small and nonsignificant.

The small but significant positive correlations between preexperimental expectations and response to test suggestions confirms previous results presented by Melei and Hilgard (1964), Dermen and London (1965), and Barber and Calverley (1966). Since previous studies had not attempted to relate preexperimental expectations to reports of unusual experiences or to testimony of having been hypnotized, the small but significant positive correlations obtained in the present study between these variables constitute new findings. These data suggest the following proposition which is relevant to a theory of hypnotic behavior:

Proposition 7. The degree to which Ss participating in a hypnotic experiment respond to test suggestions, report unusual experiences, and testify that they were hypnotized is related, in part, to their preexperimental

TABLE 3
INTERCORRELATIONS AMONG FIVE VARIABLES (PREEXPERIMENTAL EXPECTATIONS AND
FOUR DIMENSIONS OF "HYPNOTIC" BEHAVIOR)

Behavioral dimensions	Trancelike appearance	Response to test suggestions	Reports of unusual experiences	Testimony of having been hypnotized
Preexperimental expectations of hypnotic depth	.10 (.21)	.33 (.40)	.23 (.43)	.49 (.45)
Trancelike appearance		.48 (.31)	.42 (.32)	.42 (.28)
Response to test suggestions			.63 (.46)	.75 (.57)
Reports of unusual experiences				.67 (.51)

Note.—In each cell the correlations for the hypnotic-induction group is listed first and the correlation for the place-yourself-in-hypnosis group is listed next in parentheses. Correlations of .27 and above are significant at the .05 level.

expectations as to how deeply they will be hypnotized.

Antecedent Variables Determining Subjects' Testimony of Having Been Hypnotized

Situations commonly labeled as hypnotic typically include many denotable antecedent variables. Which of the antecedent variables are important and which irrelevant in determining Ss' testimony that they were hypnotized? The data obtained in the present experiment throw some light on this question.

First, Table 3 shows that, for both experimental groups, Ss' preexperimental expectations played a role in determining whether they would testify postexperimentally that they were hypnotized. Second, the final column of Table 3 shows that, for both experimental groups, testimony of having been hypnotized was positively correlated with each of the other three dimensions of hypnotic behavior—trancelike appearance, response to test suggestions, and reports of unusual experiences.

However, of the Ss who obtained above average scores on two or all three of the other behavioral dimensions, only half rated themselves as having been deeply hypnotized. Why did half of these Ss rate themselves as not deeply hypnotized and the other half rate themselves as deeply hypnotized? An answer to this question was provided by the following data. Immediately after Ss had rated their "depth" they were asked, "What basis did you use to judge your depth of hypnosis?" The Ss who had obtained above average scores on two or all three of the other behavioral dimensions (trancelike appearance, response to test suggestions, and reports of unusual experiences) and had also rated their depth as "deep" typically stated that they judged their level as deep because they had responded positively to the test suggestions (for hand lowering, finger lock, hallucination, eye catalepsy, etc.). Apparently, these Ss perceived "deep hypnosis" as involving high suggestibility. The remaining Ss, who rated their depth as "light" or "medium" even though they had obtained average scores on at least two of the three remaining behavioral dimensions, gave one or more of the follow-

ing reasons why they believed they had not entered deep hypnosis: they were aware of what they were doing, were aware of their surroundings, could think of extraneous things, could hear extraneous sounds, or did not have complete amnesia for the session. Apparently, these Ss did not rate their level as deep because their experiences did not dovetail in all respects with their preconceptions of what deep hypnosis is supposed to involve.

A recent study by Barber, Dalal, and Calverley (1968) also isolated an additional antecedent variable that is related to Ss' testimony of having been hypnotized. The latter study demonstrated that Ss' self-ratings of their hypnotic depth are influenced by the E's explicit or implicit statements that he believed they were or were not hypnotized.

In summary, the data of the present investigation, taken together with the data recently presented by Barber et al. (1968), indicate the following proposition:

Proposition 8. Whether Ss who have participated in a hypnotic experiment testify postexperimentally that they were hypnotized to a deep, medium, light, or less than light level depends on at least four sets of antecedent variables: (a) Ss' preexperimental conceptions of what hypnosis is supposed to involve; (b) their preexperimental expectations as to how deeply they will be hypnotized; (c) the degree to which, during the experiment they responded to test suggestions, manifested a trancelike appearance, and reported that they had unusual experiences; and (d) whether E stated or implied that he thought they were or were not hypnotized.

The authors believe that the data presented in the preceding sections have the following implications: (a) Many denotable antecedent variables appear to play a role in determining whether Ss will testify that they were or were not hypnotized; (b) much further work is needed to delineate the main effects and the interactive effects of all the important antecedent variables that influence Ss' testimony; and (c) it remains to be determined if Ss' testimony that they were or were not hypnotized is also related to another antecedent

factor—namely, whether or not they actually experienced an altered state of consciousness (hypnosis or hypnotic trance).

Interrelations of Four Behavioral Dimensions

It has been noted that there were small positive correlations between Ss' preexperimental expectations and each of the four sets of dependent variables (dimensions of hypnotic behavior) and also between one of the four dimensions (testimony of having been hypnotized) and the remaining three dimensions. Now look at the relationships among all four behavioral dimensions. Table 3, which presents the correlations for the two experimental groups, shows that there were significant positive intercorrelations (ranging from .28 to .75) among each of the four sets of dependent variables.⁷ These data indicate a general proposition that could well be included in a general theory of hypnotic behavior.

Proposition 9. In a hypnotic situation, Ss who are responsive on one of the dimensions of hypnotic behavior (trancelike appearance, response to test suggestions, reports of unusual experiences, or testimony of having been hypnotized) also tend to be responsive on the other three dimensions.

Although the correlations among the four dimensions are statistically significant, the small to moderate magnitudes of the correlation coefficients (which typically account for about one-quarter of the variance and in no case account for more than 56% of the variance) indicate that a substantial proportion of Ss who obtained high scores on one of the dimensions did not obtain high scores on the other dimensions. This outcome, indicating that response on each of the behavioral dimensions is at least somewhat independent of response on the other dimensions, has the fol-

lowing implications: It may be fruitful, at the present time, for behavioral scientists (a) to distinguish the four behavioral dimensions conceptually and to treat each experimentally as a separate dependent variable, (b) to use cautiously such terms as hypnotized, hypnotic state, and trance which indiscriminately refer to, and are themselves indiscriminately inferred from, all four dimensions (Barber, 1964, 1969; Chaves, 1968), and (c) to delineate experimentally the critical antecedent variables that determine high and low performance on each of the four dimensions.

REFERENCES

- BARBER, T. X. "Hypnosis" as a causal variable in present-day psychology: A critical analysis. *Psychological Reports*, 1964, 14, 839-842.
- BARBER, T. X. Measuring "hypnotic-like" suggestibility with and without "hypnotic induction"; psychometric properties, norms, and variables influencing response to the Barber Suggestibility Scale (BSS). *Psychological Reports*, 1965, 16, 809-844.
- BARBER, T. X. *Hypnosis: A scientific approach*. Princeton, N. J.: Van Nostrand, 1969.
- BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Experimental evaluation of Hull's postulate that hypnotic susceptibility is a habit phenomenon. *Journal of Personality*, 1966, 34, 417-423.
- BARBER, T. X., DALAL, A. S., & CALVERLEY, D. S. The subjective reports of hypnotic subjects. *American Journal of Clinical Hypnosis*, 1968, 11, 74-88.
- BOUCHER, R. G., & HILGARD, E. R. Volunteer bias in hypnotic experimentation. *American Journal of Clinical Hypnosis*, 1962, 5, 49-51.
- CHAVES, J. F. Hypnosis reconceptualized: An overview of Barber's theoretical and experimental work. *Psychological Reports*, 1968, 22, 587-608.
- DERMEN, D., & LONDON, P. Correlates of hypnotic susceptibility. *Journal of Consulting Psychology*, 1965, 29, 537-545.
- ERICKSON, M. H. Deep hypnosis and its induction. In L. M. LeCron (Ed.), *Experimental hypnosis*. New York: Macmillan, 1948.
- ERICKSON, M. H., HERSELMAN, S., & SECTER, I. I. *The practical applications of medical and dental hypnosis*. New York: Julian Press, 1961.
- GILL, M. M., & BRENNAN, M. *Hypnosis and related states*. New York: International Universities Press, 1959.
- HILGARD, E. R., & TART, C. T. Responsiveness to suggestions following waking and imagination instructions and following induction of hypnosis. *Journal of Abnormal Psychology*, 1966, 71, 196-208.

⁷ Although the intercorrelations among the behavioral dimensions were significant for both the hypnotic-induction and the place-yourself-in-hypnosis group, they were not significant for the control group. With respect to the control group, the correlations were as follows: -.07, between trancelike appearance and reports of unusual experiences; -.06, between trancelike appearance and testimony of having been hypnotized; and .17, between reports of unusual experiences and testimony of having been hypnotized.

- KUBIE, L. S., & MAROOLIN, S. The process of hypnosis and the nature of the hypnotic state. *American Journal of Psychiatry*, 1944, 100, 611-622.
- LONDON, P. Subject characteristics in hypnosis research. I. A survey of experience, interest, and opinion. *International Journal of Clinical and Experimental Hypnosis*, 1961, 9, 151-161.
- LONDON, P. *The Children's Hypnotic Susceptibility Scale*. Palo Alto, Calif.: Consulting Psychologists Press, 1963.
- MELEI, J. P., & HILGARD, E. R. Attitudes toward hypnosis, self-predictions, and hypnotic susceptibility. *International Journal of Clinical and Experimental Hypnosis*, 1964, 12, 99-108.
- ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.
- PATTIE, F. A. Methods of induction, susceptibility of subjects, and criteria of hypnosis. In R. M. Dorcus (Ed.), *Hypnosis and its therapeutic applications*. New York: McGraw-Hill, 1956.
- SHOR, R. E., & ORNE, E. C. *The Harvard Group Scale of Hypnotic Susceptibility. Form A*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- TART, C. T., & HILGARD, E. R. Responsiveness to suggestions under "hypnosis" and "waking-imagination" conditions: A methodological observation. *International Journal of Clinical and Experimental Hypnosis*, 1966, 14, 247-256.
- WEITZENHOFFER, A. M. *General techniques of hypnosis*. New York: Grune & Stratton, 1957.
- WEITZENHOFFER, A. M. The nature of hypnosis: Part II. *American Journal of Clinical Hypnosis*, 1963, 6, 40-72.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale: Forms A and B*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale: Form C*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.

(Received October 20, 1967)

EFFECT OF PRAISE AND CENSURE ON THE PERFORMANCE OF SCHIZOPHRENICS¹

LOUISE IRWIN² AND K. EDWARD RENNER³

University of Pennsylvania

Poor premorbid schizophrenics were shown to be primarily motivated to avoid censure and good premorbid to be relatively more sensitive to praise. It was demonstrated that when censure was used on a task with only two possible responses, poor premorbid perform better than good premorbid. The opposite relationship was demonstrated for the effects of praise, to which the good premorbid were more responsive. On a task with many responses the praise or censure gave little information as to the correct response. On this task, censure was demonstrated to be relatively more disruptive to the poor premorbid than praise, but the opposite to hold for the good premorbid. Comparisons with neutral conditions and normal Ss were also made.

Data are not in agreement with regard to the effects of praise and censure on the performance of schizophrenics. In some studies, it was found that schizophrenics performed more poorly under censure than under neutral conditions (Alvarez, 1957; Dunham, 1959; Garnezy, 1952; Neiditch, 1963; Webb, 1955). Other investigators have found, however, that schizophrenics performed better under censure than under neutral conditions (Atkinson & Robinson, 1961; Cavanaugh, 1958; Cavanaugh, Cohen, & Lang, 1960; Olson, 1958). In some studies schizophrenics when praised did not show an increment in performance in comparison to their performance under neutral conditions (e.g., Cavanaugh et al., 1960); although in a few studies their performance did improve under praise (e.g., D'Alessio & Spence, 1963).

Review of the literature suggests that the level of premorbid adjustment and the nature of the task can account for the differential performance shown by schizophrenics under praise and censure. Garnezy (1965) postulated that good premorbid schizophrenics are less sensitive to censure than poor premorbid

schizophrenics; also, on the basis of life history data (Phillips, 1953), one would expect to find differences in sensitivity to praise and censure between good and poor premorbid schizophrenics. Poor premorbid begin to avoid social contact in early childhood, while good premorbid continue to participate in social activities throughout their premorbid period. Thus, the premorbid adjustment of poor premorbid is predominately one of avoiding social contacts, while that of good premorbid is patterned around an approach to others.

The nature of the task may also be important. Atkinson and Robinson (1961) and Johannsen (1961) postulate that the schizophrenic's reaction to censure is dependent on the number of ways available to avoid censure in the situation. However, the findings of some studies are contradictory to this interpretation (Fischer, 1963; Goodstein, Guertin, & Blackburn, 1961; Losen, 1961). The important factor may not be the number of ways available to avoid censure, but rather the amount of information given S concerning the means to avoid it. Thus, in a simple situation where only two responses are possible, censure should provide complete information about what is the correct response and the most efficient way to avoid censure is to give the correct response. However, on a complex task where many responses are possible, censure would not give information and should produce task irrelevant responses.

The present study was designed to test

¹This article is based in part on a PhD dissertation submitted by the first author to the University of Pennsylvania in partial fulfillment of the requirements for the PhD degree.

²Now at Delaware State Hospital, Wilmington, Delaware.

³Request for reprints should be sent to K. Edward Renner who is now at the Department of Psychology, University of Illinois, Urbana, Illinois 61801.

explicitly this interpretation of past research. Specifically, on a simple task censure should improve performance for poor premorbid schizophrenics who are primarily motivated to avoid censure, while good premorbid should show some sensitivity to praise under these conditions. On the other hand, in a complex task where the praise or censure does not provide information as to the correct response, task irrelevant responses should appear and the schizophrenic deficit be found. Under these conditions the poor premorbid schizophrenics should show a greater decrement in performance under censure than good premorbid, but the good premorbid should be relatively more sensitive to the effects of praise. In analysis of variance terms this is a prediction of a triple interaction of Level of Premorbid Adjustment \times Difficulty of the Task \times Effects of Praise or Censure. This interpretation is consistent with the reported empirical results on schizophrenic performance and resolves most of the apparent contradictions.

METHOD

Subjects

The Ss were 36 normals, and 36 good and 36 poor premorbid schizophrenics. All Ss were selected in terms of the following criterion: (a) males between 20 and 55 yr. of age; (b) with WAIS vocabulary scores of at least 6; (c) with no known pathology of the central nervous system; (d) with no known history of alcoholism; and (e) not undergoing shock therapy (although some Ss had probably received shock treatment at some time in the past). The normal Ss were chosen from the population of male employees of Embreeville State Hospital, who had no history of a neuropsychiatric illness. Good and poor premorbid Ss were chosen from diagnosed schizophrenics on the basis of their scores on the Phillips Scale of Premorbid Adjustment in Schizophrenia (Phillips, 1953). Following Garnezy and Rodnick (1959) the ratings of premorbid adjustment were based on an interview rather than case history data. The range of possible scores was 0-30, a high score indicating poor premorbid adjustment. A score of 18 and above or 13 and below were required in order to be included in the sample. The reliability of the ratings was determined by having an independent rating made on a tape recording of 20 of the interviews; this yielded a correlation coefficient of .86.

Half of the Ss of the normal, good, and poor premorbid groups were assigned to the low-complexity condition, and half to the high-complexity

condition. All Ss were run under a neutral, a praise, and a censure condition.

The 36 good premorbid consisted of 27 paranoid, 3 catatonic, 2 simple, 2 chronic undifferentiated, and 3 acute undifferentiated schizophrenics. The 36 poor premorbid included 22 paranoid, 3 catatonic, 3 hebephrenic, 3 simple, and 5 chronic undifferentiated schizophrenics. The mean age at first hospitalization of the good premorbid was 32.1 yr. and for the poor premorbid 23.8 yr. The mean length of total hospitalization was 30 mo. for the good premorbid and 70 mo. for the poor premorbid. Age of first hospitalization and length of hospitalization were significantly different ($p < .01$), as is typically the case for good and bad premorbid schizophrenics (e.g., McInnis & Ullmann, 1967).

Task

The stimuli consisted of 144 five-letter words, which occur with equal and high frequency in the English language (Thorndike & Lorge, 1944). Each word was typed on a white cardboard square which was placed behind a cellophane window of a 2-in. square coin holder. Either two or six of these coin holders, each displaying 1 word in each window, were mounted on the blank pages of a scrap book to form six books each containing eight pages. The S was required to guess and remember which of the words on each of the eight pages had been arbitrarily chosen by E to be the correct response for that page. Complexity was varied by presenting Ss in the low-complexity condition with 2 words so that they had to eliminate only one alternative in order to learn the correct response. In the high-complexity condition, however, each page contained 6 words. In this condition, therefore, S had to learn which of the six alternatives was the correct response for each of the eight pages. Two pilot studies were run to obtain three sets of words of comparable difficulty for each of the complexity conditions. The order of presentation of these three sets of words was counterbalanced for each of the subgroups so that each set appeared in the neutral, praise, and censure positions an equal number of times.

Reinforcement

In the neutral condition, Ss were given no social reinforcement. In order for them to obtain feedback about the correctness of their responses, the coin holders were mounted on the pages by means of hinges of cellophane tape. The correct word appeared in the space behind the coin holder as well as on its face, while the spaces behind the holders displaying the incorrect words were blank. Therefore, after S announced his guess, he could check behind the holder displaying the word to see if it was correct.

In the praise and censure conditions, Ss were given information about the correctness of their guess by means of two boxes, each measuring $17 \times 8 \times 8$ in., placed side by side on a table at which S

was seated and located 1½ ft. in front of and facing S. The boxes could be illuminated with the word "right" appearing on the panel of one and the word "wrong" on the panel of the other. A failure to respond was treated as an incorrect response. In the instructions for the neutral conditions, Ss were given only the task requirements. However, the instructions stressed the importance of being right in the praise condition and of not being wrong in the censure condition.

Procedure

All data were collected by the first author who conversed with each S for a few minutes to permit S to become acclimated to the experimental situation. In the first session all Ss were run under the neutral condition. In a second session, 2 days after the first, Ss learned one set of words under praise and one set of words under censure, the two orders counterbalanced throughout the design. There was a 15-min. break between the two sets and during this period Ss were given either a cup of coffee or glass of coke depending on their preference.

The Ss were allowed to spend 10 sec. on each page. After they announced their guess and checked to see if they were correct in the neutral condition they were then told to turn to the next page. If Ss failed to make a guess by 7 sec. they were told that their time was nearly up. Each run through the eight pages of the book was designated as 1 trial. The Ss were run until they met the criterion of 1 perfect trial or until they had completed 8 trials on the simple task or 15 trials on the complex task.

RESULTS

Performance of Schizophrenic Subjects

The major interest focuses on the relative performance of the schizophrenic Ss under praise and censure on the simple and complex task. The data are illustrated in Figure 1. Three terms of the analysis of variance reached significance. Of principle interest, however, is the predicted triple interaction; $F = 4.58$, $df = 1/68$, $p < .05$. The interaction resulted from the good premorbid performing relatively better than the poor premorbid on a simple task under praise but with the poor premorbid performing relatively better under censure. However, on the complex task the poor premorbid schizophrenics made relatively fewer errors under praise than under censure, as compared with the good premorbid. The nature of the triple interaction is in the predicted direction.

Overall, the good premorbid schizophrenics performed better than the poor premorbid

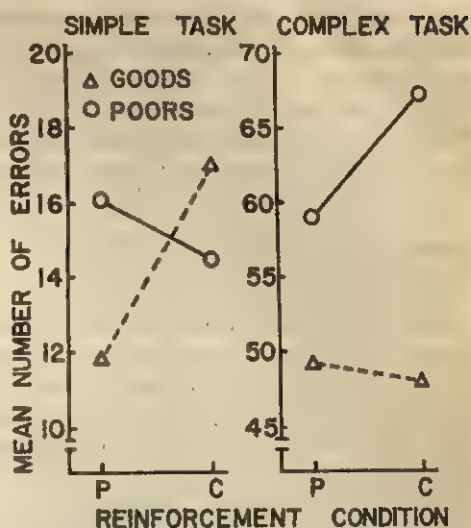


FIG. 1. Mean number of errors for good and poor premorbid schizophrenics under praise (P) and censure (C) on the simple (left panel) and the complex (right panel) task.

schizophrenics, $F = 4.87$, $df = 1/68$, $p < .05$. The fact that this better performance occurred primarily on the complex task, as shown in Figure 1, did not yield a significant interaction between level of premorbid adjustment and task complexity, $F = 3.86$, $df = 1/68$, $p < .10$.

The third significant term resulted from the fact that more errors were made on the complex than the simple task as would be expected, $F = 142.78$, $df = 1/68$, $p < .001$.

Comparison with Normal Subjects

Of secondary interest is a comparison of the performance of schizophrenics with the performance of normals which allows an evaluation of the relative deficit under the various conditions. On the simple task there was no significant schizophrenic deficit between the normals and poors, $F = 1.77$, $df = 1/34$, or between the normals and the goods, $F = .04$, $df = 1/34$. In fact the goods actually made fewer errors under praise than the normals and the poors fewer errors under censure than normals as may be seen in Table 1, although averaged over the conditions the normal Ss were slightly superior. The picture is radically changed, however, on the complex task where

the schizophrenic deficit was expected to show itself. The normals performed better than the good premorbid, $F = 6.25$, $df = 1/34$, $p < .05$; and the special sensitivity of the poor premorbid to censure was reflected in the large deficit they showed on the complex task, $F = 25.58$, $df = 1/34$, $p < .001$.

Comparison of Neutral with Social Reinforcement

Of additional interest is the within-Ss comparison of the neutral condition with praise and censure for the three groups. The normals did not change their performance on the simple task when either praise or censure were introduced with $t(17) = .155$ and $t(17) = .600$, respectively. This may be contrasted with the fact that poor premorbid on the simple task made significantly less errors under censure than under neutral conditions with $t(17) = 1.854$, $p < .05$, one-tailed, but not under praise where $t(17) = 1.390$, $p = ns$. On the other hand, good premorbid performed significantly better under praise than under neutral conditions with $t(17) = 1.949$, $p < .05$, one-tailed, but not under censure where $t(17) = .073$, $p = ns$.

On the complex task, however, there was a uniform finding of better performance under both praise and censure than under neutral conditions for all groups. This uniformity of better performance under conditions of social reinforcement than under the neutral condi-

tion was contrary to expectations. Thus, on the complex test the schizophrenic decrement in performance, which was quite large in some instances, occurred only in comparison with normal Ss but not in comparison with the schizophrenic's own performance under the neutral condition. Thus, even on the complex task the social reinforcement facilitated performance over neutral conditions.

DISCUSSION

The results confirmed the principal expectations. Normal Ss reacted similarly to praise and censure, whereas schizophrenics responded differentially to these reinforcers. The performance of schizophrenics under praise and censure was dependent both upon their premorbid adjustment and on the difficulty of the task. Poor premorbid showed facilitation on the simple task and interference on the complex task under censure in comparison to their performance under praise. This performance difference reflected their presumed strong motivation to avoid censure, which in the case of the simple task could be accomplished most efficiently by giving the correct answer, but on the complex task, by schizophrenic withdrawal which produced the large decrement under censure relative to the effects of praise. On the other hand, good premorbid showed the opposite relationship, reflecting their responsiveness to praise and their lower sensitivity to censure.

Similar findings for a simple task were reported by McInnis and Ullmann (1967) who obtained an interaction between responsiveness to positive and negative reinforcement by short- and long-term hospitalized schizophrenics. The direction of the differences corresponded perfectly with the finding of greater responsiveness to censure by poor premorbid and to praise by good premorbid on a simple task. As noted previously, in both studies the two schizophrenic groups differed on both length of hospitalization and premorbid history. The analysis focusing on premorbid history is supported by the differential sensitivity obtained in both studies. McInnis and Ullmann's primary interest was in length of hospitalization. They expected, and found, short-term hospitaliza-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF ERRORS UNDER PRAISE AND CENSURE MADE BY THE NORMALS AND BY THE GOOD AND BAD PREMORBID SCHIZOPHRENIC SUBJECTS ON THE SIMPLE AND COMPLEX TASK

Task	Reinforcement condition					
	Neutral		Praise		Censure	
	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
Simple						
Normals	13.3	9.5	13.0	8.7	15.0	8.7
Good premorbid	17.3	10.7	11.9	7.6	17.0	11.6
Poor premorbid	20.4	10.8	16.1	9.3	14.5	9.4
Complex						
Normals	46.5	12.2	37.9	12.0	40.2	15.6
Good premorbid	58.7	19.0	49.2	19.9	48.5	19.7
Poor premorbid	76.8	21.0	59.0	20.1	67.4	22.6

tion to be related to greater sensitivity to both positive and negative reinforcement. In this study, no such main effect was found for the poor-good premorbid dimension on the simple task. This failure to confirm their results is not surprising, however, for in this study good premorbid had an average hospitalization which was nine times longer than the short-term schizophrenics in the McInnis and Ullmann study. Nearly all of the Ss in this study would be regarded as long term by their standards. Thus, long-term hospitalization may reduce responsiveness to both positive and negative reinforcement, but differential sensitivity seems to be a function of premorbid history.

The second problem of theoretical interest is the evaluation of the absolute level of performance by comparing the performance of schizophrenics with that of normals. This is a between-Ss comparison rather than the within-Ss comparison used in the preceding section to evaluate the *relative* effects of praise and censure. This second comparison offers further support for the hypothesis that schizophrenic deficit is of a motivational nature. When incentives appropriate to the motivations of the two types of schizophrenics were used on a task where the motivation would result in a correct answer the difference between normals and schizophrenics was minimal with the schizophrenic performance slightly better. On the other hand, there were large differences between the performance of normal and schizophrenic Ss on the complex task when the motivation was more likely to produce schizophrenic withdrawal rather than the correct answer. The consideration of task difficulty along with premorbid history can thus account for the fact that in previous studies schizophrenics sometimes did better and sometimes worse than normals under praise and censure.

A final observation of interest is the fact that schizophrenics did not show a decrement in performance under praise and censure on the complex task relative to their *own* performance under neutral conditions. This finding was unexpected. It must be remembered, however, that all Ss were run in the neutral condition first and were treated kindly by E.

It may be that these Ss had more confidence in E and were more at ease during the second session, or simply that a practice effect produced the absolute differences. The neutral condition was always presented first in order to insure a more sensitive within-Ss comparison of the relative effects of praise and censure. This particular design was chosen because the neutral versus reinforcement comparison was regarded as the least important. The results tentatively support the suggestion that under experimental conditions which put the schizophrenic patient at ease he will respond to social reinforcement. Such a finding is consistent with a motivational interpretation of schizophrenic deficit, which exists in an absolute sense in comparison to normals, but not in comparison to his own performance under neutral conditions. Such an hypothesis is difficult to evaluate, however without an independent measure of whether E has enlisted the cooperation and trust of the schizophrenic patient. Such trust may well be a necessary condition for E to become effective as a reinforcing agent for the schizophrenic. These observations must remain suggestive, however, for the design was not appropriate or intended to give an unambiguous answer on these issues.

REFERENCES

- ALVAREZ, R. R. A comparison of the preferences of schizophrenics and normal subjects for rewarded and punished stimuli. Unpublished doctoral dissertation, Duke University, 1957.
- ATKINSON, R. L., & ROBINSON, N. M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, 62, 322-326.
- CAVANAUGH, D. K. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *Journal of Abnormal and Social Psychology*, 1958, 57, 8-12.
- CAVANAUGH, D. K., COHEN, W., & LANG, P. J. The effect of "social censure" and "social approval" on the psychomotor performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1960, 60, 213-218.
- D'ALESSIO, G. R., & SPENCE, J. T. Schizophrenic deficit and its relation to social motivation. *Journal of Abnormal and Social Psychology*, 1963, 66, 390-393.

- DUNHAM, R. M. Sensitivity of schizophrenics to parental censure. Unpublished doctoral dissertation, Duke University, 1959.
- FISCHER, E. Task performance of chronic schizophrenics as a function of verbal evaluation and social proximity. *Journal of Clinical Psychology*, 1963, 19, 176-178.
- GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality*, 1952, 21, 253-276.
- GARMEZY, N. The prediction of performance in schizophrenia. In P. H. Hoch & J. Zubin (Eds.), *Schizophrenia*. New York: Grune & Stratton, 1965.
- GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease*, 1959, 129, 450-466.
- GOODSTEIN, L. D., GUERTIN, W. J., & BLACKBURN, H. L. Effects of social motivational variables on choice reaction time of schizophrenics. *Journal of Abnormal and Social Psychology*, 1961, 62, 24-27.
- JOHANNSSEN, W. J. Responsiveness of chronic schizophrenics and normals to social and nonsocial feedback. *Journal of Abnormal and Social Psychology*, 1961, 62, 106-113.
- LOSEN, S. M. The differential effects of censure on the problem-solving behavior of schizophrenic and normal subjects. *Journal of Personality*, 1961, 29, 258-272.
- MCINNIS, T. L., & ULLMANN, L. Positive and negative reinforcement with short- and long-term hospitalized schizophrenics in a probability learning situation. *Journal of Abnormal Psychology*, 1967, 72, 157-162.
- NEIDTCH, S. J. Differential response to failure in hospital and nonhospital groups. *Journal of Abnormal and Social Psychology*, 1963, 66, 449-453.
- OLSON, G. W. Failure and subsequent performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1958, 57, 310-314.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.
- WEBB, W. W. Conceptual ability of schizophrenics as a function of threat of failure. *Journal of Abnormal and Social Psychology*, 1955, 50, 221-224.

(Received October 23, 1967)

VISUAL MONITORING PERFORMANCE IN THREE SUBGROUPS OF MALE DELINQUENTS¹

JAMES BURDEANE ORRIS²

University of Illinois

A questionnaire measuring psychopathic, neurotic, and subcultural delinquency was administered to 268 male delinquents. Three experimental groups of 15 boys each were formed from Ss who scored above the mean on one scale and below the mean on the other two. It was predicted that the more psychopathic Ss would show a greater performance decrement on a vigilance task. A significant effect was found between experimental groups and a significant decrement over trials for both signals detected and reaction times; psychopathic Ss performed consistently poorer than the other two groups.

The present study is an investigation into the nature of the psychopathy dimension of delinquency identified in previous research (e.g., Peterson, Quay, & Cameron, 1959; Quay, 1964). It was hypothesized that more psychopathic (P) Ss would show a performance decrement relative to more neurotic (N) and more subcultural (S) Ss on a task requiring continuous attention. This prediction was made on the basis of a theoretical approach (Quay, 1965) to psychopathy which postulates that psychopaths are pathological stimulation seekers and are highly susceptible to boredom, thus likely to exhibit a decrement in monitoring efficiency because of a tendency to look elsewhere for added stimulation.

The classical vigilance experiment (e.g., Mackworth, 1961) in which S has to watch for and report transient signals presented against a background of neutral signals was chosen as the experimental task.

METHOD

Subjects. A factor analytically derived questionnaire measuring the P, N, and S dimensions of delinquency (Quay & Peterson, 1964) was used to select Ss. From a sample of 268 institutionalized male delinquents three experimental groups of 15 Ss each were formed from boys who scored above the

mean on one scale and below the mean on the other two.

Apparatus. The apparatus was a visual monitoring display consisting of a voltmeter with the dial reversed so that the pointer would be seen against a white background. The pointer was deflected from resting position to middial position at a regular rate of one deflection per second. The critical signal was a deflection of the pointer approximately three-fourths of an inch greater than the standard halfway deflection. The Ss responded to the signal by pressing a hand-held microswitch. The experimental room was dark except for a small light used to illuminate the display. A fan was used to produce a constant noise which masked any extraneous sounds.

In an adjoining room was a timer which started automatically whenever a signal was presented to S and which stopped whenever S made a response. The timer could be stopped by E if S did not respond and was reset by E after each signal.

Procedure. Each S participated in a 100-min. practice session in the morning and an identical 100-min. criterion session in the afternoon; sessions were divided into five 20-min. trials for statistical analysis. Ten signals were presented during each trial; intersignal intervals from 20 sec. to 240 sec. were arranged in a different random order for each of the five trials.

Response measures were both number of signals detected and reaction times. If S responded within 4 sec., E recorded his reaction time; otherwise a "miss" was recorded.

RESULTS

Figure 1 presents data for the criterion session with number of signals detected per trial as the performance measure. There was a significant effect for the three experimental groups, $F = 4.36$, $df = 2/42$, $p < .05$, and a significant effect over trials, $F = 8.40$, $df = 4/168$, $p < .01$. The interaction between groups and trials was not significant, nor

¹ This paper is based on a master's thesis at the University of Illinois under the direction of Herbert C. Quay and Jack A. Adams. Appreciation is expressed to Anthony P. Travisano, Superintendent, and the staff of the Iowa Training School for Boys.

² Requests for reprints should be sent to Herbert C. Quay, Temple University, College of Education, Philadelphia, Pennsylvania 19122.

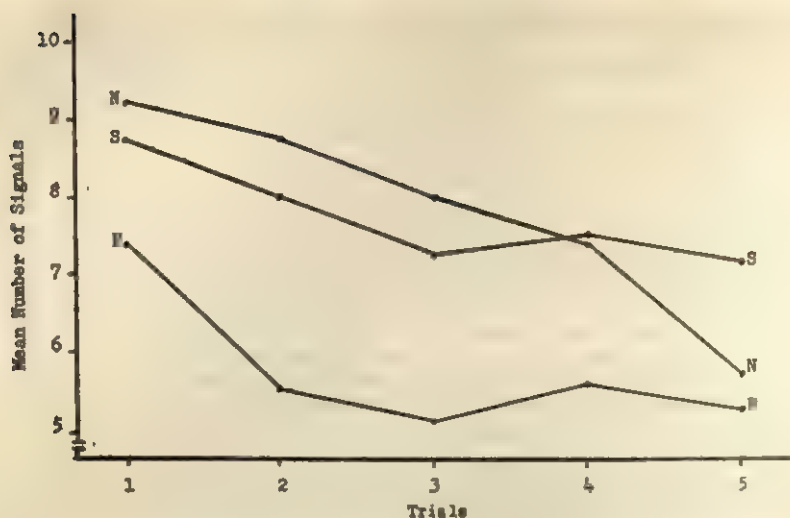


FIG. 1. Mean number of signals detected per trial during criterion session.

were the groups significantly different on the first trial.

Criterion sessions using reciprocal latency as the performance measure are presented in Figure 2. The effect for groups, $F = 5.03$, $df = 2/42$, $p < .05$, and for trials, $F = 11.60$, $df = 4/168$, $p < .01$, were both significant, while the interaction between groups and trials was not significant; the groups were

significantly different on the first trial, $F = 3.39$, $df = 2/42$, $p < .05$. Since the N and S Ss were essentially identical, this difference was due to the poorer performance of the P Ss.

DISCUSSION

While, as predicted, the more psychopathic Ss exhibited a significant vigilance decrement,

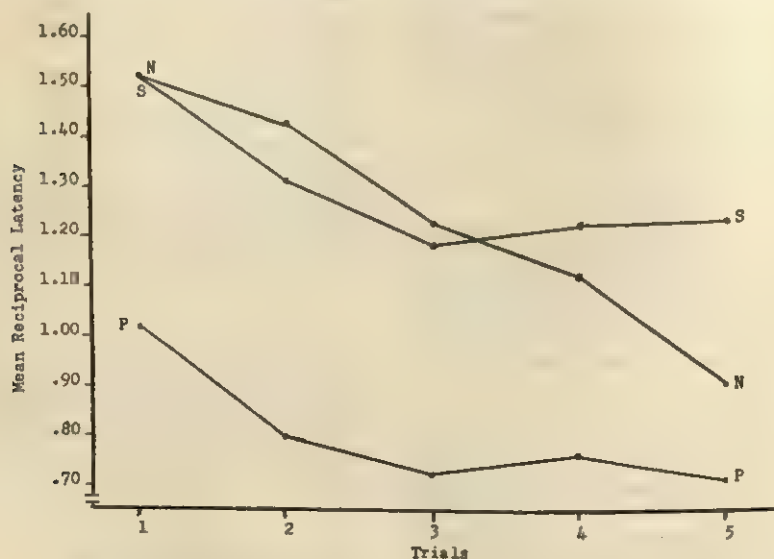


FIG. 2. Mean reciprocal latency per trial during criterion session.

the lack of a significant interaction between groups and trials indicated that the groups did not adapt or habituate to the setting at significantly different rates. Examination of the performance curves reveals that the mean scores for the P group were below the other groups for all trials. The difference between groups was significant even on the first trial of the criterion session for the reciprocal latency scores, but not for number of signals detected. Thus the vigilance decrement occurred early and was maintained throughout.

Clinically it was observed that the P group engaged in more self-stimulation, such as singing and talking, during the course of the experiment than did the other two groups. Such observations should be quantified in subsequent research.

REFERENCES

- MACKWORTH, N. H. Researches on the measurement of human performance. In H. W. Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems*. New York: Dover, 1961.
- PETERSON, D. R., QUAY, H. C., & CAMERON, G. R. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. *Journal of Consulting Psychology*, 1959, 23, 395-399.
- QUAY, H. C. Personality dimensions in delinquent males as inferred from the factor analysis of behavior ratings. *Journal of Research in Crime and Delinquency*, 1964, 1, 33-37.
- QUAY, H. C. Psychopathic personality as pathological stimulation seeking. *American Journal of Psychiatry*, 1965, 122, 180-183.
- QUAY, H. C., & PETERSON, D. R. The questionnaire measurement of personality dimensions associated with juvenile delinquency. Unpublished manuscript, University of Illinois, 1964.

(Received October 27, 1967)

BEHAVIORAL AND PHYSIOLOGICAL CHANGES DURING PROLONGED IMMOBILIZATION PLUS PERCEPTUAL DEPRIVATION¹

JOHN P. ZUBEK,² L. BAYER, S. MILSTEIN, AND JEAN MARY SHEPHARD

University of Manitoba

The Ss who successfully completed 1 wk. of immobilization plus perceptual deprivation (IPD group) showed a greater slowing of occipital EEG activity, and a poorer performance on a battery of intellectual and perceptual-motor tests than did Ss exposed to a similar duration of either immobilization (I group) or a recumbent control condition (RC group). During the 1-wk. period, the IPD group also showed a significant increase in urinary excretion of noradrenaline, but not of adrenaline, relative to the I and RC groups. No significant differences were observed on behavioral measures of subjective stress and mood.

In a series of experiments from the Manitoba laboratory (Zubek, 1968; Zubek & MacNeill, 1966; Zubek & Wilgosh, 1963), it has been shown that Ss whose level of kinesthetic-proprioceptive stimulation was reduced by immobilization but who otherwise were exposed to a normal and varied sensory environment for a week exhibited a variety of cognitive and perceptual-motor deficits, unusual subjective phenomena, a slowing of the electrical activity of the brain, a reduced urinary excretion of catecholamines (adrenaline and noradrenaline), and no changes in subjective stress and mood. However, a comparison of the results of the immobilized Ss with those of control Ss revealed that most of the effects were associated with either the recumbent position or with a combination of recumbency and restricted motor activity. Since only a few of the changes resulted solely from immobilization, the purpose of the present study was to determine whether increasing the severity of the experimental condition by the addi-

tional restriction of visual and auditory stimulation (perceptual deprivation) produces greater behavioral and physiological changes than immobilization alone.

METHOD

Subjects

A group of 30 male university students were placed individually in a "coffinlike" box, located inside a soundproofed room, for a prescribed period of 1 wk. The S was requested to lie in the box with his head in a padded head holder and his feet in a set of V-shaped restraining holders. The arm, legs, and trunk were immobilized by straps fastened to the base of the box and kept sufficiently loose so as not to interfere with circulation. To minimize possible pain and physical discomfort, Ss were unstrapped periodically, for example, for 15 min. at mealtimes and for approximately 70 min. in the afternoon (for testing and washroom purposes). They were also unstrapped for 9 hr. during the night but were not allowed to stand up or sit up (for details of the immobilization box and procedure see Zubek & MacNeill, 1966). In addition to the restriction on motor activity, Ss constantly wore a white face mask (except during the periodic test administrations) which permitted diffuse light (20 ftc. under the mask) but eliminated pattern vision and were exposed to white noise, somewhat above the threshold of hearing, produced by a speaker located behind S's head. The overhead lights and the white noise were kept on during the entire 1-wk. period. Conversation between S and E was kept to a minimum. At least one E was on duty at all times to take care of such needs as feeding, occasional adjustments of restraining straps, etc. A moderate degree of unavoidable social interaction existed, therefore, between S and E, more so than normally occurs in the usual perceptual deprivation experiment.

¹This research was supported by Grant MH-08748 from the National Institutes of Health, United States Public Health Service and by Grant 9425-08 from the Defence Research Board, Canada. The authors wish to express their appreciation to M. Nickerson, Department of Pharmacology, for providing the biochemical determinations and to D. Dickson, K. Fuerst, L. Nepon, D. Oleson, and W. Schutte for research assistance during the conduct of the experiment.

²Request for reprints should be sent to John P. Zubek, Department of Psychology, University of Manitoba, Winnipeg, Canada.

Because of the severity of the experimental condition, only 12 of the 30 Ss, the successfals, were able to endure the 1-wk. period. The remainder, the quitters, terminated the experiment after a mean duration of 33.4 hr.; 13 within the first day and the remaining 5 Ss by the end of the fourth day.

Measures

Intellectual battery. This battery, developed by Zubek, Sansom, and Prysiazniuk (1960), consists of 11 different tests of a primary mental ability nature, with eight equivalent forms per test, measuring numerical reasoning, arithmetic problems, abstract reasoning, verbal reasoning, verbal fluency, space relations, digit span, recall, recognition, cancellation of numbers, and dexterity. This battery was administered before and at daily intervals during the 1-wk. experimental period. Total administration time was approximately 50 min.

Perceptual-motor tests. Two tests, namely depth perception and size constancy, were administered before and after the 1-wk. period. Depth perception was measured by the Howard-Dohlman apparatus and size constancy by means of a black, equilateral triangle, of variable height, presented at a distance of 15 ft. The S was required to adjust the height of this triangle until it looked the same in size as the near one, the standard triangle (for further procedural details see Zubek & MacNeill, 1966). These two tests were employed because earlier research has indicated that neither performance measure is affected by either a week of immobilization (Zubek & MacNeill, 1966; Zubek & Wilgosh, 1963) or 1 to 2 wk. of perceptual deprivation (Zubek, 1964; Zubek, Aftanas, Hasek, Sansom, Schludermann, Wilgosh, & Winocur, 1962).

EEG activity. Electroencephalograms were taken from 9 of the 12 successful experimental Ss before and after the 1-wk. period. The records were taken at the same time of the day with precautions to eliminate any drowsiness. The tracings were taken on an Offner eight-channel, Type T, machine. Sixteen Ediswan monopod electrodes were applied to the head in International Federation 10/20 positions and two S.L.E. suction electrodes to the ears. Average, bipolar, and monopolar recordings were taken. In order to obtain a quantitative measure of EEG changes, the mean occipital lobe frequency of each S, before and after the 1-wk. period, was determined according to the method of Engel, Romano, Ferris, Webb, and Stevens (1944). This involved counting, by means of a Marshall ruler, the number of waves occurring in each of 200, 1-sec. samples of artifact-free occipital lobe tracings and then obtaining the mean frequency per second. In a few instances, 200 artifact-free samples could not be obtained and therefore a somewhat smaller sample had to be employed. EEG records were also taken from 11 of the 18 quitters before and after their premature termination of the experiment.

Biochemical and affective measures. Urine samples were taken at 24-hr. intervals for 3 days prior to

the experiment, at daily intervals during the experiment, and for 2 days after its termination resulting in twelve 24-hr. samples for each S who successfully endured the 1-wk. period. An analysis for catecholamines, known to be involved in various stressful situations (e.g., Elmadjian, Hope, & Lamson, 1957; Frankenhaeuser & Patkal, 1964; Levi, 1964), was then performed by the fluorimetric procedure of Euler and Lishajko (1961) and was expressed as total excretion of adrenaline and noradrenaline, separately, per 24-hr. period. In addition to this biochemical analysis, an appraisal was made of Ss' affective state by two paper-and-pencil tests. The first was the Subjective Stress Scale (SSS) which has been shown to differentiate well between individuals who have experienced situations inducing various degrees of stress (Berkun, Bialek, Kern, & Yogi, 1962; Myers, 1964). The SSS is a 15-item Thurstone scale, with higher scores indicating more adverse reactions. The second test, developed by Myers, Murphy, Smith, and Goffard (1966) specifically for perceptual deprivation experiments, is a measure of overall mood. It consists of a 114-item adjective checklist which S answers as applying "not at all," "somewhat or slightly," or "mostly or generally" to him during a particular time period. Both the mood test and the SSS were administered at daily intervals before, during, and after the experimental period, that is, at intervals corresponding to the 24-hr. urine samples.

Comparison Groups

The results of the 12 successful, perceptually deprived immobilized Ss (IPD group), on the biochemical and affective measures, were compared with the results of 16 immobilized (I group) and 16 recumbent control Ss (RC group) from a recent study (Zubek, 1968) employing the same coffinlike box, 1-wk. duration, and measures as the IPD group but using Ss exposed to normal visual and auditory stimulation. Furthermore, to control for possible seasonally associated changes in catecholamine excretion both studies were conducted during the same time of the year. The results of the IPD group on the remaining measures, namely intellectual, perceptual-motor, and EEGs were compared with the results of another 1-wk. study (Zubek & MacNeill, 1966) employing 32 immobilized and 32 recumbent control Ss. To control for possible differences in the level of initial ability, the preexperimental scores of the 12 successful IPD Ss, on the 11 intellectual and 2 perceptual-motor measures, were matched with the initial scores of 12 of the 32 immobilized (I group) and 12 of the 32 recumbent control Ss (RC group).

RESULTS

Intellectual battery. An analysis of variance on the scores (mean of 7 daily tests administered during the 1-wk. period) of the three groups of preexperimentally matched Ss (12

per group) was performed on each of the 11 tests of the intellectual battery. Following this analysis, the significance of the differences between the various groups of Ss was evaluated by two-tailed t tests.

Significant F values were obtained for six of the tests, namely verbal fluency ($F = 4.27$, $p < .05$) space relations ($F = 5.80$, $p < .01$), abstract reasoning ($F = 3.83$, $p < .05$), dexterity ($F = 6.32$, $p < .01$), cancellation ($F = 12.99$, $p < .001$), and recall ($F = 7.27$, $p < .01$). The performance of the IPD group was significantly worse than that of either the I or RC groups on tests of verbal fluency (p 's $< .05$), space relations ($p < .01$; $p < .05$), abstract reasoning (p 's $< .05$), and dexterity ($p < .01$; $p < .05$). None of the differences between the I and RC groups were significant. On the other hand, the performance on cancellation and recall was poorer in the IPD than in the I group ($p < .001$; $p < .01$) and in the I group than in the RC group ($p < .01$; $p < .05$). Although the F values were significant for only 6 of the 11 intellectual tests, it is important to note that the IPD group did worse than the I group on all of the remaining five tests.

No statistical analysis of the results of the 18 IPD quitters was performed since most of them terminated the experiment prior to the administration of the first intellectual battery on Day 1.

Perceptual-motor tests. The F values for the three groups of matched Ss (12 per group) on the measure of size constancy and depth perception were both significant, $F = 8.75$, $p < .001$; $F = 3.44$, $p < .05$. A t -test analysis revealed that the IPD group performed significantly worse than did either the I or RC group on size constancy (larger comparison triangles were chosen) and depth perception ($p < .01$; $p < .05$, respectively). None of the differences between the I and RC groups was significant.

EEG records. The relative change in mean occipital lobe frequency during the 1-wk. period, for each of the IPD ($N = 9$), I ($N = 12$), and RC ($N = 12$) groups was -1.42 , $-.59$, and $-.11$ cps, respectively ($F = 8.77$, $p < .001$). A t test revealed that the IPD group showed a significantly greater

decrease in mean frequency than the I group ($p < .01$) and the I group a greater decrease than the RC group ($p < .01$). An analysis was also made of the results of 11 IPD quitters receiving EEG measures before and after a mean duration of 29.5 hr. (range of 3–96 hr.). The mean decrease was $-.72$ cps, a change significantly smaller than the decrease of -1.42 cps shown by the nine, 1-wk. IPD Ss ($p = .01$). The rank-order correlation between duration and the magnitude of the EEG change of the 11 quitters was $+.74$ ($p < .01$).

Catecholamine determinations. Figure 1 (bottom) shows the 24-hr. urinary excretion of adrenaline of the 12 successful IPD, 16 I, and 16 RC Ss. Although all three groups show a decreased excretion during the 1-wk. experimental period, it is interesting to note that the excretion levels of both the IPD and I groups are consistently higher than the level of the RC group throughout the entire experimental period. However, an analysis of variance on the difference scores between the preexperimental (mean of 3 days) and "during" experimental periods (mean of 7 days) of the three groups of Ss yielded a non-significant F value.

Figure 1 (bottom) also shows the pre- and postexperimental adrenaline levels of the 18 IPD quitters. It can be seen that these levels are consistently lower than those of the three other groups on the 3 days before and the 2 days after the termination of the experiment. Despite this trend, none of the differences were significant. The results of the quitters during the experimental period are not shown in Figure 1 since most of them terminated the condition prior to the end of the first day and thus a 24-hr. urine sample was not available for analysis. Furthermore, a proportional conversion of their adrenaline values into a 24-hr. period was considered to be inappropriate because of the existence of diurnal variations in catecholamine excretion (Euler, Hellner-Bjorkman, & Orwen, 1955).

Figure 1 (top) shows the urinary excretion of noradrenaline, the neurohormone of the peripheral sympathetic system (Euler, 1956), in the successful IPD, I, and RC Ss. It can be

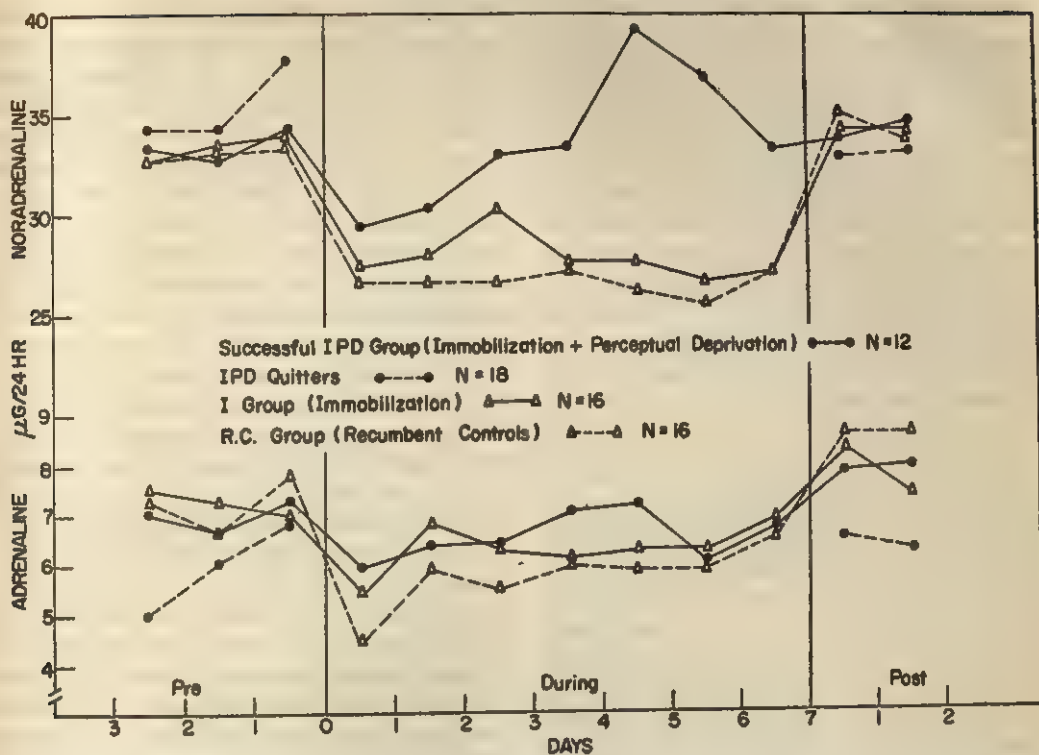


FIG. 1. The 24-hr. urinary excretion of adrenaline (bottom) and noradrenaline (top) of four groups of Ss before, during, and after a 1-wk. period. (The catecholamine levels of the IPD quitters, during the experimental period, are not shown in the figure since most of them terminated the condition prior to the end of the first day and thus a 24-hr. urine sample was not available for analysis.)

seen that both the I and RC groups show a pronounced decrease in noradrenaline excretion, and of approximately the same magnitude, throughout the entire 1-wk. period, an effect probably related to the maintenance of the recumbent position which is known to reduce the excretion of both noradrenaline and adrenaline (Euler, Luft, & Sundin, 1955; Sundin, 1958). On the other hand, the IPD group, after an initial decrease, shows a gradual increase in noradrenaline excretion reaching a peak during the fifth day. An analysis of variance on the difference scores between the preexperimental and during experimental periods of the three groups of Ss yielded a significant F value ($F = 3.88$, $p < .05$). Furthermore, the IPD group showed a significantly greater mean increase in noradrenaline excretion relative to that of either the I or RC group (p 's $< .05$). The differ-

ence between the I and RC groups was not significant.

A comparison of the pre- and postexperimental noradrenaline levels of the 18 IPD quitters with those of the three successful groups revealed no significant differences. It is interesting to note, however, the presence in the quitters of a trend toward increased noradrenaline excretion on the day prior to the experiment and relative to the preceding 2 days, a trend also seen for adrenaline. This increased catecholamine excretion which has also been reported for other studies, for example, in perceptual deprivation quitters a day prior to isolation (Zubek & Schutte, 1966) and before centrifuge (Goodall & Berman, 1960) or drug studies (Elmadjian et al., 1957) is believed to be indicative of apprehension about experimental participation. Surprisingly, however, no such trend

toward increased catecholamine excretion appears to be present in the three groups of Ss who successfully endured the 1-wk. period.

Affective measures. The results of the two paper-and-pencil tests, which were administered at the time intervals corresponding to those of the 24-hr. catecholamine determinations, revealed that the IPD group showed a consistently higher subjective stress score and a more negative mood, relative to both the I and RC groups, during the 1-wk. experimental period. However, on neither of the two affective measures were any of the differences statistically significant.

DISCUSSION

The results of this study have demonstrated that increasing the severity of the immobilization condition by the additional restriction of visual and auditory stimulation (perceptual deprivation) produces greater behavioral impairments than does immobilization alone. This was shown by the presence of a significant deficit in depth perception and size constancy in the IPD but not in the I group. Furthermore, since it has been reported that neither of these perceptual-motor measures is affected by a week of either sensory or perceptual deprivation in an isolation chamber (Zubek et al., 1962; Zubek, Pushkar, Sansom, & Gowing, 1961), it would appear that an impairment of depth perception and size constancy will only occur after a severe and prolonged restriction of visual, auditory, and kinesthetic-proprioceptive stimulation. Significant impairments in the IPD relative to the I group were also observed on 6 of the 11 tests of the intellectual battery and a poorer performance, though not statistically significant, on the remaining 5 tests. Accompanying these greater behavioral impairments, the IPD group showed a significantly greater decrease in mean occipital lobe frequency (-1.42 cps), after the 1-wk. period, and relative to that of either the I or RC group ($-.59$, $-.11$ cps, respectively). In addition, since a high positive correlation (.74) was obtained between the pre-post difference scores of the IPD quitters and their duration, the results suggest the presence of a greater slowing of EEG activity with increasing dura-

tions of the IPD condition. This decrease in frequency of 1.42 cps in the 1-wk. IPD group, it is important to note, is greater than that obtained in an earlier study (Zubek & Welch, 1963) employing a week of either sensory ($-.85$ cps) or perceptual isolation (-1.21 cps) with no immobilization. This difference in magnitude of EEG slowing may account, for example, for the deficit in depth perception and size constancy in the IPD condition but no change in performance on these two measures after the same duration of sensory and perceptual isolation (Zubek et al. 1962; Zubek, Pushkar, Sansom, & Gowing, 1961). Some experimental support for this hypothesis has been reported by Pollack (1963) who noted that various psychiatric treatments which induced different degrees of EEG slowing were accompanied by deficits on a variety of intellectual and perceptual tests. Furthermore, the magnitude of these deficits was related to the degree of induced EEG slowing. In view of this association, it is possible that the greater intellectual impairments observed in the IPD relative to the I group may also be related to the difference in EEG slowing occurring between these two groups of Ss.

The results of the biochemical determinations revealed no significant differences in adrenaline excretion in the three groups of 1-wk. Ss during the experimental period. However, for noradrenaline the excretion level of the IPD group increased significantly relative to both the I and RC groups. Since no significant differences on either catecholamine measure were obtained between the I and RC group, or in an earlier 1-wk. study (Zubek & Schutte, 1966) comparing perceptually isolated and recumbent control Ss, the results suggest that an increase in noradrenaline but not in adrenaline will only occur after a severe restriction of visual, auditory, and kinesthetic-proprioceptive stimulation.

Although the IPD group showed a significant increase in noradrenaline, this biochemical effect was not accompanied by any significant changes on the subjective stress and mood scales, a disparity sometimes noted in other studies relating biochemical and behavioral measures of stress (e.g., Cofer & Appley,

1964; Lazarus, 1964). Furthermore, this increase in noradrenaline, particularly noticeable during the last 3 days of the experimental period, does not appear to be related to Ss' greater degree of outwardly expressed anger ("anger-out") which several groups of investigators (e.g., Elmadjian et al., 1957; Funkenstein, King, & Drolette, 1957) have claimed is associated with an increased production of noradrenaline. An examination of the appropriate items in the 114-item mood scale, administered daily, revealed no evidence for an increased incidence of anger-out or aggressive emotional reactions in the IPD group relative to the I and RC groups. An observation of their daily behavior also failed to detect any differences in this respect. A more likely explanation of the elevated noradrenaline level is that it is a reflection of an accumulating, growing emotional tension, of a general nature, developing over time.

The differential pattern of results obtained in the IPD group, namely an increased excretion of noradrenaline but not of adrenaline has also been observed in other "stress" conditions, for example, during water immersion (Graveline, Balke, McKenzie, & Hartman, 1961), performance of insoluble tests (Frankenhaeuser & Patkai, 1964), and in hockey coaches during a game (Elmadjian et al., 1957). Although the reasons for this differential pattern are not clear, such results might be expected in view of the different site of production of adrenaline and noradrenaline (Euler, 1956).

Finally, an interesting trend was observed in the IPD quitters, namely, a consistent tendency to possess a lower base-line level of adrenaline, relative to the successful 1-wk. Ss, during the 3 days before and 2 days after the experiment. This trend is significant in the light of two earlier 1-wk. studies indicating a similar, but reliable, base-line difference in adrenaline between successful and unsuccessful perceptual isolation Ss (Zubek & Schutte, 1966) and a difference, bordering on statistical significance, between successful Ss and quitters undergoing prolonged immobilization (Zubek, 1968). One possible explanation for the nonsignificant results in the IPD quitters may be the greater severity

of their experimental condition as indicated by a quitting rate of 60% versus approximately 40% for the Ss in the other two conditions. Despite this nonsignificant trend, it is intriguing that the base-line levels of adrenaline but not of noradrenaline, of the three types of quitters, should all be low.

REFERENCES

- BERKUN, M. M., BIALEK, H. M., KERN, R. P., & YOGI, K. Experimental studies of psychological stress in men. *Psychological Monographs*, 1962, 76(15, Whole No. 534).
- COFER, C. N., & APLEY, M. H. *Motivation theory and research*. New York: Wiley, 1964.
- ELMAJIAN, F., HOPE, J. M., & LAMSON, E. T. Excretion of epinephrine and norepinephrine in various emotional states. *Journal of Clinical Endocrinology*, 1957, 17, 608-620.
- ENGEL, G. L., ROMANO, J., FERRIS, E. B., WEBB, J. P., & STEVENS, C. D. A simple method of determining frequency spectrums in the electroencephalogram. *Archives of Neurology and Psychiatry*, 1944, 51, 134-146.
- EULER, U. S. VON. *Noradrenaline*. Springfield, Ill.: Charles C Thomas, 1956.
- EULER, U. S. VON, HELLNER-BJORKMAN, S., & ORWEN, I. Diurnal variations in the excretion of free and conjugated noradrenaline and adrenaline in urine from healthy subjects. *Acta Physiologica Scandinavica*, 1955, 33, 10-16.
- EULER, U. S. VON, & LISHAJKO, F. Improved technique for the fluorimetric estimation of catecholamines. *Acta Physiologica Scandinavica*, 1961, 51, 348-355.
- EULER, U. S. VON, LUFT, R., & SUNDIN, T. The urinary excretion of noradrenaline and adrenaline in healthy subjects during recumbency and standing. *Acta Physiologica Scandinavica*, 1955, 34, 169-174.
- FRANKENHAEUSER, M., & PATEAL, P. Interindividual differences in catecholamine excretion during stress. *Report of the Psychological Laboratory*, No. 178, 1964, University of Stockholm.
- FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. E. *Mastery of stress*. Cambridge, Mass.: Harvard University Press, 1957.
- GOODALL, M., & BERMAN, M. L. Urinary output of adrenaline, noradrenaline, and 3-methoxy-4-hydroxymandelic acid following centrifugation and anticipation of centrifugation. *Journal of Clinical Investigation*, 1960, 39, 1533-1538.
- GRAVELINE, D. E., BALKE, B., MCKENZIE, R. E., & HARTMAN, B. Psychobiologic effects of water-immersion-induced hypodynamics. *Aerospace Medicine*, 1961, 32, 387-400.
- LAZARUS, R. S. A laboratory approach to the dynamics of psychological stress. *American Psychologist*, 1964, 19, 400-411.

- LEVI, L. The urinary output of adrenaline and noradrenaline during different experimentally induced pleasant and unpleasant states. A summary. *Journal of Psychosomatic Research*, 1964, 8, 197-198.
- MYERS, T. I. Studies of stressfulness of sensory deprivation and confinement. Paper presented at Eastern Psychological Association, Philadelphia, April 1964.
- MYERS, T. I., MURPHY, D. B., SMITH, S., & GOFFARD, S. J. Experimental studies of sensory deprivation and social isolation. Technical Report 66-8, June 1966, Human Resources Research Office, George Washington University.
- POLLACK, M. Psychological changes with induced EEG slowing. *American Psychologist*, 1963, 18, 443. (Abstract)
- SUNDIN, T. The effect of body posture on the urinary excretion of adrenaline and noradrenaline. *Acta Medica Scandinavica*, 1958, 1-59 (Suppl. 336).
- ZUBEK, J. P. Behavioral and EEG changes after 14 days of perceptual deprivation. *Psychonomic Science*, 1964, 1, 57-58.
- ZUBEK, J. P. Urinary excretion of adrenaline and noradrenaline during prolonged immobilization. *Journal of Abnormal Psychology*, 1968, 73, 223-225.
- ZUBEK, J. P., AFTANAS, M., HASEK, J., SANSOM, W., SCHLUDERMANN, E., WILGOSH, L., & WINOCUR, G. Intellectual and perceptual changes during prolonged perceptual deprivation: Low illumination and noise level. *Perceptual and Motor Skills*, 1962, 15, 171-198.
- ZUBEK, J. P., & MACNEILL, M. Effects of immobilization: Behavioral and EEG changes. *Canadian Journal of Psychology*, 1966, 20, 316-336.
- ZUBEK, J. P., PUSHKAR, D., SANSOM, W., & GOWING, J. Perceptual changes after prolonged sensory isolation (darkness and silence). *Canadian Journal of Psychology*, 1961, 15, 83-100.
- ZUBEK, J. P., SANSOM, W., & PRYSIAZNIUK, A. Intellectual changes during prolonged perceptual isolation (darkness and silence). *Canadian Journal of Psychology*, 1960, 14, 233-243.
- ZUBEK, J. P., & SCHUTTE, W. Urinary excretion of adrenaline and noradrenaline during prolonged perceptual deprivation. *Journal of Abnormal Psychology*, 1966, 71, 328-334.
- ZUBEK, J. P., & WELCH, G. Electroencephalographic changes after prolonged sensory and perceptual deprivation. *Science*, 1963, 139, 1209-1210.
- ZUBEK, J. P., & WILGOSH, L. Prolonged immobilization of the body: Changes in performance and in the electroencephalogram. *Science*, 1963, 140, 306-308.

(Received November 1, 1967)

EFFECTS OF TWO INDEPENDENT STRESSES UPON AUTONOMIC RESPONDING¹

JAMES H. GEER²

AND

KENNETH KLEIN³

State University of New York at Stony Brook

University of Minnesota

In an attempt to determine whether the GSR and cardiac response to "aversive" photographs contained both orienting and emotional components, 40 Ss were shown either photographs of dead bodies or photographs of live individuals projected upside down. In addition, half of the Ss expected, but never received, shocks. The data were interpreted to indicate that in addition to the orienting response, an emotional component was present in the response to the aversive photographs. The effect of shock threat was to increase reactivity to all stimuli. Analysis revealed that the GSR was most responsive to the content of the stimuli while the cardiac response was most sensitive to the shock threat.

In a recent study (Geer, 1966) Ss selected as being frightened of spiders but not of snakes yielded larger GSRs when shown pictures of spiders than when shown pictures of snakes. In that study it was noted that there were two possible explanations for the increase in GSR responding to the aversive pictures. One explanation was that the stimuli aroused a negative emotion; the second was that the stimuli aroused an increased orienting response (OR) that did not necessarily contain elements of negative emotion. The present study was designed to test the alternative explanations in an altered experimental design. Different stimuli were used, and Ss were not selected for their fearfulness of those stimuli. The aversive stimuli used in this study were color photographs of victims of violent death, which were considered noxious, but not painful.

A second major interest of this study is the attempt to assess the effects of an independent threat or stress upon responding to

aversive and nonaversive stimuli. The independent threat was produced by telling Ss that they would receive electrical shock at points throughout the experiment. It was predicted that the effect of the shock threat would be to raise responsiveness to both the aversive and the nonaversive stimuli.

The final area of interest that is examined in this study relates to the autonomic systems that are being measured. Both cardiac activity and GSRs were collected and examined. Cardiac activity was considered to be of particular importance since Graham and Clifton (1966) suggest that different components of the cardiac response reflect orienting and defensive responses. If that suggestion is accurate, it should be possible to further evaluate the OR versus negative emotion interpretations.

METHOD

The Ss were 40 female undergraduates enrolled in introductory psychology at the University of Pennsylvania. Four of these Ss were eliminated on the basis of recording and procedural errors. All Ss were met by E who conducted them to a sound and electrically shielded room. They were seated in a reclining chair that was placed directly in front of a reverse projection window upon which the stimuli were delivered. Electrodes for recording cardiac and GSR activity were attached, and S was given general instructions to move as little as possible and to look at each picture. All Ss were told that they would be viewing pictures of both live individuals and pictures of dead bodies and that E would be recording their physiological reaction to those stimuli. In addition one-half of the Ss, randomly selected, were told that from time to time "electrical stimulation" would be applied through two electrodes placed on

¹This investigation was supported by Public Health Service Research Grant MH12301-02 from the National Institutes of Health. The data were collected while the authors were at the University of Pennsylvania. Portions of this paper were read at the Eastern Psychological Association Convention, April 1967, Boston, Massachusetts.

²Requests for reprints should be sent to James H. Geer, Department of Psychology, State University of New York, Stony Brook, New York 11790.

³The authors wish to thank the Medical Examiners Office, city of Philadelphia, Pennsylvania for its cooperation in the procurement of the photos of dead bodies.

the left hand. The remaining Ss were told that this second set of electrodes were for measuring an aspect of skin reactivity. In fact, no shocks were given any Ss. A brief rest period of approximately 5 min. followed while recording equipment was checked and calibrated.

All Ss were shown 10 color photographs of male undergraduates. The photographs, taken against a brick wall background, were of models bare from the waist up. For each S the same 10 photographs were shown; however, the order was randomized. The presentation of 10 photographs was used to permit partial habituation of responses to "neutral" stimuli. Following Trials 1-10 with these habituation stimuli, the second experimental manipulation was introduced. On Trials 11-15 one-half of the Ss, randomly selected and equally divided between shock-threat and no-shock-threat conditions, saw photographs of more undergraduate males, but the photographs were projected upside down. This was the nonaversive condition. The remaining Ss on Trials 11-15 were shown color photographs of male victims of sudden or violent death. This was the aversive condition. Each photograph was of the upper torso of the individual and in most instances the autopsy table top served as the background. Blood and other evidence of trauma were present in the photos. The order of stimuli on Trials 11-15 was randomized for each S.

Photographs were presented at intertrial intervals of from 50 to 70 sec. with a mean intertrial interval of 60 sec. Intertrial intervals were controlled by a programmed timer, and the sequence of intervals varied for each S. Each photo was projected by a Kodak Carousel projector for 5 sec.; the stimulus duration was controlled by an electronic timer. GSRs and cardiac rates were recorded on a Beckman Model RB dynograph using that instrument's couplers as the response sensing devices. Beckman electrodes and electrode paste were used for all data collection.

RESULTS

A GSR was defined as a measurable decrease in skin resistance that began within 1 to 5 sec. after stimulus onset. The measure of the response on a given trial was the log of the change in conductance. The formula used was $\left(\frac{1}{R_2} - \frac{1}{R_1}\right) \times 10^9 + 1.0$ where R_1 is the resistance in ohms at the start of the inflection and R_2 is the resistance at the point of maximum deflection. For each S the mean GSR on Trials 8, 9, and 10 was computed and those data were subjected to a 2×2 analysis of variance. The same analysis was performed on the mean log of the basal conductance (R_1) for Trials 8, 9, and 10. These analyses produced no significant results, indicating that neither basal skin resistance nor

reactivity to the last three habituation stimuli were significantly affected by instructions. There were no detectable differences between groups assigned to the aversive and non-aversive conditions.

Each S's mean GSR for Trials 8, 9, and 10 was computed and subtracted from each GSR S gave on Trials 11 through 15. The principle dependent variable used for the GSR was the mean of those difference scores. This measure has the advantage of subtracting each S's general GSR reactivity from the response to the experimental stimuli. For a S to obtain a positive score she had, on the average, to respond with greater amplitude to test than to the mean of the last three habituation stimuli. Table 1 contains the group means for the GSR dependent variable. A 2×2 analysis of variance of those data yielded a highly significant ($F = 14.33$, $df = 1/32$, $p < .001$) aversive versus nonaversive stimulus effect while neither the Shock Instructions nor the Aversive versus Nonaversive Stimuli \times Shock Instructions interaction approached significance. Since a check of homogeneity of variance determined that that assumption for the analysis of variance was violated, several nonparametric analyses were conducted to verify the analysis reported above. The median test indicated ($\chi^2 = 12.00$, $df = 3$, $p < .01$) that the four groups differed in reactivity to stimuli. The same statistic indicated that the aversive group was considerably more reactive ($\chi^2 = 11.8$, $df = 1$, $p < .001$). A similar analysis for the shock-threat conditions did not approach statistical significance. The Man-Whitney U test was used to compare individual groups. The shock-threat-aversive group differed significantly ($p < .002$) from both nonaversive groups. No other individual comparisons were significant.

TABLE 1

MEAN CHANGE IN THE LOG OF CHANGE
IN CONDUCTANCE SCORES

	Dead bodies	Upside down photos	Total
Shock threat	.944	.112	1.056
No shock threat	.550	.064	.614
Total	1.494	.176	

Note.—Mean of Trials 8, 9, and 10 from Trials 11, 12, 13, 14, and 15.

These analyses were interpreted to indicate that GSR reactivity was greater to dead bodies as stimuli than to upside down photographs, and that the difference was more pronounced under shock threat.

According to Sokolov (1963) the OR habituates while the defensive response either does not or does so at a reduced rate. This suggests that frequency analyses may be useful in discriminating between the two types of responses. Thus, a 2×2 analysis of variance of the number of GSRs to the five experimental stimuli was performed. The results were similar to the amplitude data: aversive versus nonaversive stimulus effect was highly significant ($F = 22.48$, $df = 1/32$, $p < .001$) while neither shock instructions nor the interaction term approach statistical significance. These results support the interpretation that the aversive stimuli, pictures of dead bodies, elicited more than a simple OR to stimulus change.

The heart rate data were analyzed by analyses of variance as the homogeneity of variance assumptions were met with these data. An analysis was performed on the mean heart rate of the two interpulse intervals that preceded the stimulus on Trial 11. As with the comparable GSR data, no effects approaching statistical significance were found. Apparently instructions did not affect basal heart rate levels. Following the suggestion by Lang and Hnatiow (1962) and Graham and Clifton (1966) the cardiac response was considered to be made up of an accelerative and a decelerative component. The accelerative component of the biphasic cardiac response was measured by finding the difference in heart rate between the mean of two interpulse intervals that immediately preceded stimulus onset and the fastest beat that occurred within 1-5 sec. after stimulus onset. As with the GSR data, the mean of the accelerative component was computed for Trials 8, 9, and 10 and this value was subtracted from each accelerative component on Trials 11, 12, 13, 14, and 15. Table 2 contains the group means for the accelerative component. An analysis of variance was performed upon the mean of each individual's difference scores. A significant shock instructions effect ($F = 5.12$, $df = 1/32$, $p < .05$) was found, indicating that threat of shock was associated

TABLE 2
MEAN CHANGE IN CARDIAC ACCELERATION
(BEAT PER MINUTE)

	Dead bodies	Upside down photos	Total
Shock threat	3.82	.76	4.58
No shock threat	-.26	-4.60	-4.86
Total	3.56	-3.84	

Note.—Mean of Trials 8, 9, and 10 from Trials 11, 12, 13, 14, and 15.

with an increase in the accelerative component. A decrease in the accelerative component was found to stimuli presented under no shock threat. The aversive versus nonaversive stimulus effect just missed accepted levels of statistical significance ($F = 3.14$, $df = 1/32$, $p < .10$, $p > .05$). These results are taken as a strong indication that shock threat yielded increased cardiac acceleration. The interaction term did not approach statistical significance. A frequency analysis of the number of increases in the cardiac accelerative component yielded the same finding of a significant effect of shock instructions ($F = 6.26$, $df = 1/32$, $p < .05$); neither aversive versus nonaversive stimulus effects nor the interaction effects approach significance.

The decelerative component of the cardiac response was measured by subtracting the slowest beat subsequent to the previously noted accelerative beat from the prestimulus measure. The decelerative beat was restricted to the 15 sec. that followed stimulus onset. As with the accelerative component, individual reactivity was calculated by determining the mean of the decelerative component on Trials 8, 9, and 10. The dependent variable was the difference, for each S, between the decelerative component on Trials 11 through 15 and the mean of Trials, 8, 9, and 10. Table 3 contains the group means for the difference in decelerative components. A 2×2 analysis of variance was performed upon the data and the results closely parallel that for the accelerative component. The effect of shock instructions ($F = 8.62$, $df = 1/32$, $p < .01$) was significant. Table 3 indicates that the decelerative component was larger under shock threat than under conditions of no threat. The aversive versus non-

TABLE 3
MEAN CHANGE IN CARDIAC DECELERATION
(BEATS PER MINUTE)

	Dead bodies	Upside down photos	Total
Shock threat	3.48	1.60	5.08
No shock threat	-.72	-4.26	-4.98
Total	2.76	-2.66	

Note.—Mean of Trials 8, 9, and 10 from Trials 11, 12, 13, 14, and 15.

aversive stimulus effect again was not significant but was in the direction of increased reactivity (decelerative) to dead bodies. The interaction effect was also negligible in this instance. A frequency analysis of the number of increases in the decelerative component yielded similar results: shock instructions were significant ($F = 18.8$, $df = 1/32$, $p < .001$) while aversive versus nonaversive stimulus effects and the interaction did not attain statistical significance. The heart rate data were interpreted to indicate that both components of the biphasic cardiac response were sensitive to the shock threat, and less sensitive to content of the stimuli.

DISCUSSION

The GSR was apparently more affected by the aversive photos of dead bodies than by upside down photographs, as indicated by larger and more frequent GSRs among Ss in the aversive group. The data also suggested that this effect was enhanced by shock threat. The cardiac data are consistent with the GSR data though they do not actually achieve accepted levels of statistical significance. For both the accerelative and decelerative components, dead bodies tended to elicit greater cardiac reactivity and this was particularly true under conditions of shock threat.

The use of upside down photographs was designed to provide nonaversive stimuli that were more dissimilar from habituation stimuli than were the aversive stimuli. It should be noted that stimulus difference is used here only in reference to the absolute physical properties of the stimuli. It may be argued that the upside down photographs differed from the habituation photographs only in a trivial sense, but, it is indeed just that point which this study attempts to illustrate.

Sokolov's (1963) theory concerning the OR states that an OR is elicited by stimulus change and that OR amplitude increases with increased stimulus change. Further Sokolov (1963) is explicit in referring only to physical properties of the stimulus. "The term [neuronal model] suggests that the nervous system produces an exact model of the properties of external objects acting on the sense organs [p. 286]."

In an attempt to verify that the physical characteristics of the stimuli differ as proposed above, judges were asked to rate stimulus differences on the dimension of physical characteristics. Instructions included explicit directions to ignore startle or shock value and to respond as if they (the judges) were viewing photographs of unknown objects. Judges, without exception, rated the physical characteristics of the aversive stimuli to be more similar to the habituation stimuli than were the nonaversive stimuli. It should be noted that all judges pointed out that this was true only if one totally ignored stimulus content. The point to be made is that when considering stimuli presented to an S, one cannot rely only upon sense receptor stimulation but must also consider prior experiences of S.

The augmented GSR to the aversive stimuli is viewed as indicating that more than a simple OR to stimulus change was aroused. This view is supported by the analysis that revealed reduced habituation of the GSR to the aversive stimuli. While what this extra something is cannot be stated with assurance, it seems reasonable to infer that it reflects the negative emotion elicited by the aversive stimuli. What cannot be determined is whether this response to the perception of the content of the stimuli results in an augmented OR, or a more direct reflection of the negative emotion as such. If Sokolov's (1963) definition of the defensive response as being "closely connected with the sensation of pain [p. 15]" is maintained, then the stimuli used in this study did not elicit the defensive response. It is clear that the content of the stimulus is an important variable, and the content may elicit more than an OR to stimulus change. Perhaps the effect of certain stimulus content is similar to what occurs when a stimulus change acquires signal properties. When a stimulus or stimulus change

acquires signal properties, as in the case of a conditioned stimulus, OR habituation is inhibited. Only further research can clarify these issues.

Consider now the effect of the independent stress, the threat of shock. Several interesting things may be noted concerning this manipulation. First, there is no evidence that general arousal, as indexed by either basal levels of GSR or cardiac rate, was effected by the shock threat. The effect of the shock threat was upon reactivity to novel stimuli. It was as if *S* under shock threat was primed to respond to changes in his environment. It should be noted that in no case did the interaction of instructions and stimuli approach statistical significance. That is, independent stress seems to effect reactivity in general and does not act to differentially increase reactivity to more arousing stimuli.

The cardiac response was separated into accelerative and decelerative components as suggested by Graham and Clifton (1966). There were no detectable differences between the components of the response. Both components indicated that the cardiac response to novel stimuli is increased when *S* is under the threat of shock. The GSR data, while not attaining statistical significance, were in the same direction as the cardiac data and provided further evidence that the effect of shock threat was to augment reactivity to all novel stimuli, aversive or not.

The differential sensitivities of the cardiac and GSR are of considerable interest. First, it should be noted that the two systems did not yield contradictory findings. The results indicated that the GSR was most sensitive to the content of the stimuli while the cardiac response was most sensitive to the shock threat. Why this should be the case is not clear. It is, however, consistent with Lacey's (1950) notion of relative independence of the components of the autonomic system. The data are also consistent with the finding by Dykman, Reese, Galbrecht, and Thomasson (1959) that in normal humans the GSR is more stimulus bound than cardiac or respiration reactions, and the interpretation of Goldstein and Acker (1967) that skin resistance measures index, "the readiness of the organism to attend to, or be alerted by, external stimuli [p. 28]." At a speculative level, it

might be suggested that the cardiac response is best manipulated by potentially important physical threats to the individual. Photographs of dead bodies provide no actual threat to the physical integrity of the individual.

The finding that cardiac acceleration and deceleration were consistent is not in complete accord with Graham and Clifton's (1966) suggestion that they index different responses. These authors suggested that the accelerative component may be associated with the defensive or perhaps startle reflex, and the decelerative component with the OR. If one accepts Sokolov's (1963) contention that the defensive reflex is elicited through painful stimuli, then the data are not supportive of Graham and Clifton's position since no actual pain was aroused. If, however, a broader interpretation of defensive reflex is employed, one that includes potentially painful stimuli, or startle stimuli, then it could be argued that stimulus change elicited the OR and the shock threat elicited the defensive reflex. Thus, in the present study both aspects of the cardiac response may have been affected. Until further research clarifies the startle, orienting, and defensive components of the cardiac response, any interpretation must remain tentative.

REFERENCES

- DYKMAN, R. A., REESE, R. G., GALBRECHT, C. R., & THOMASSON, P. G. Psychophysiological reactions to novel stimuli: Measurement, adaptation, and relationship of psychological and physiological variables in the normal human. *Annals of the New York Academy of Science*, 1959, 79, 43-107.
- GEER, J. H. Fear and autonomic arousal. *Journal of Abnormal Psychology*, 1966, 71, 253-255.
- GOLDSTEIN, M. J., & ACKER, C. W. Psychophysical reactions to films by chronic schizophrenics: II. Individual differences in resting levels and reactivity. *Journal of Abnormal Psychology*, 1967, 72, 23-29.
- GRAHAM, F. K., & CLIFTON, R. K. Heart rate change as a component of the orienting response. *Psychological Bulletin*, 1966, 65, 305-320.
- LACEY, J. I. Individual differences in somatic response patterns. *Journal of Comparative and Physiological Psychology*, 1950, 43, 338-350.
- LANG, P. J., & HNATIOW, M. Stimulus repetition and the heart rate response. *Journal of Comparative and Physiological Psychology*, 1962, 55, 781-785.
- SOKOLOV, E. N. *Perception and the conditioned reflex*. New York: Pergamon Press, 1963.

(Received November 3, 1967)

SCHIZOPHRENIA AND THE SCANNING COGNITIVE CONTROL: A REEVALUATION

T. MCKINNON¹ AND G. SINGER²

University of Sydney

In recent studies it has been inferred from performance on perceptual tasks, that paranoid schizophrenics showed extensive scanning and nonparanoid schizophrenics showed minimal scanning. Eye movements during free-search and size-judgment tasks were recorded for paranoid and nonparanoid schizophrenics, psychotic depressives, and control Ss. Medicated patients showed lower eye-movement rate and fixation rate than nonmedicated patients and controls but there were no significant differences due to schizophrenia. There was a significant interaction between patient type and medication for size-judgment means, but all such differences were reduced to nonsignificance by training. However, even after training, patients showed greater variability in size judgments than control Ss. The results failed to support the notion of extensive or minimal scanning behavior of schizophrenics. Errors by the psychotic patients in making size adjustments were not correlated with scanning behavior.

The concept of scanning—the extensiveness with which stimuli are sampled when attending to a sensory or perceptual field—was introduced by Gardner, Holzman, Klein, Linton, and Spencer (1959) to refer to certain stylistic consistencies in perceptual and cognitive functioning of normal Ss. A factor from perceptual judgment measures loading on size estimations and size constancy judgments was interpreted as reflecting scanning although no direct visual scanning measures were included in the study. Gardner and Long (1962a, 1962b) obtained direct measures of visual scanning by recording eye movements during size-estimation tasks. The various eye-movement measures for different conditions of the size-matching task involving minor changes in the standard stimulus were positively correlated. This provided some evidence of individual consistencies in direct measures of visual scanning although the correlations between the eye-movement measures and the size judgments were low.

Silverman (1964a) employed the concept of scanning to summarize and explain results of a number of perception experiments in-

volving schizophrenics, and as the basis for a general theory of schizophrenic behavior.

Silverman (1964b) tested 26 nonparanoid and 17 paranoid male schizophrenics all receiving tranquilizer medication on the standard Gardner et al. (1959) size-estimation task; the patients were asked to adjust a circular patch of light to the same size as each of three hand-held disks. Nonparanoid schizophrenics showed consistent overestimation and paranoid schizophrenics showed underestimation. These differences were significant at the .01 level or better.

On the basis of the Gardner and Long studies, Silverman interpreted the differences in size estimation as reflecting differences in scanning behavior without obtaining any direct measures of visual scanning. It was concluded that paranoid schizophrenics showed extensive scanning—continually alternating fixations, looking back and forth—and nonparanoid schizophrenics showed minimal scanning—maintaining fixation with little eye movement.

The aim of the present study was to test directly the postulated differences in scanning behavior between paranoid and nonparanoid schizophrenics and normal Ss.

Definition of Visual Scanning

In the Gardner and Long (1962b) study in which the scanning process was defined directly by eye-movement measures, the size-

¹ Now at the University of New South Wales.

² The cooperation of the New South Wales Department of Public Health and the hospital staff and patients is acknowledged. Requests for reprints should be sent to George Singer who is now at the School of Behavioral Sciences, MacQuarie University, North Ryde, New South Wales, 2113, Australia.

estimation procedure involved *S* difference in judgment time. The eye-movement measures (particularly those found to correlate significantly with size judgments) were confounded with judgment time. In this case a high scanner was a *S* who repeatedly looked back and forth from the standard to the comparison stimulus and showed long judgment time. However, in other tasks (autokinesis, size constancy, size judgment) from which Silverman (1964a) inferred scanning differences, judgment time or exposure time was mostly held constant. In these cases Silverman implies that differences in the eye-movement *rate* define the scanning process. The problem arises that groups of extensive and minimal scanners defined by the different criteria (*a*) total number of eye movements in a variable judgment time task and (*b*) rate of eye movements in a constant exposure task may be unrelated or negatively related.

Before the present study was undertaken, the stability and generality of individual consistencies in measures of rate and number of eye movements of normal *Ss* were investigated. The corneal photography technique was used to record eye movements in a wide range of conditions including perceptual judgments (McKinnon, 1966).

This study showed stability of eye-movement measures over time and the generality of individual consistencies in saccadic *eye-movement rate* over a broad range of experimental conditions. Visual scanning rate was unrelated to "functional efficiency" of the visual system defined by fixation steadiness, tracking efficiency, and binocular coordination. There was restricted evidence for individual consistencies in adjustment time and related measures of the number of saccadic eye movements for size- and brightness-adjustment tasks. However, none of the eye-movement measures was significantly correlated with size judgments. For experimentally independent conditions, visual scanning rate was *consistently negatively correlated* with time taken to make perceptual adjustments.

These findings indicate difficulties in inferring scanning differences from perceptual judgments alone and the need to specify objective criteria of scanning.

In the present study the perceptual and scanning behavior (defined by the number and rate of eye movements) of medicated- and nonmedicated-schizophrenic groups, psychiatric, and normal controls was examined for free-search and size-judgment conditions.

It is important to specify conditions under which differences in size judgments or scanning behavior of schizophrenics and normal or psychiatric control *Ss* are expected to occur and whether these can be eliminated.

Stannard, Singer, and Over (1966) reported that differences between the performance of schizophrenic and normal *Ss* on a size-constancy task under normal motivation conditions were eliminated when behavior was manipulated using impersonal reinforcement in an operant procedure.

In the present study a training procedure in the size estimation task involving verbal feedback from *E* was employed to attempt to eliminate any deficit shown by patients in size judgments.

If size underestimation and overestimation are associated with extensive and minimal scanning, respectively, then if training eliminates size-judgment deficit, it could be expected to alter scanning patterns.

METHOD

Subjects. Six groups of 14 *Ss*, each including seven males and seven females, were tested. The six groups were: (*a*) nonparanoid schizophrenic, medicated; (*b*) nonparanoid schizophrenic, nonmedicated; (*c*) paranoid schizophrenic, medicated; (*d*) paranoid schizophrenic, nonmedicated; (*e*) psychotic depressive, nonmedicated; and (*f*) psychiatric nurses (control).

The psychiatric patients were selected from admission wards at two metropolitan psychiatric hospitals. Clinical diagnosis was made by the psychiatrist treating the patient and doubtful cases were excluded. All patients were tested within 4 wk. of admission with not more than a total of 2 yr. of hospitalization. There was an age range of 20–50 yr. (mean age of the nonparanoid schizophrenic was 34.5 yr., $SD = 9.8$ yr.; paranoid schizophrenic: $M = 38.3$ yr., $SD = 11.0$ yr.; psychotic depressive: $M = 44.6$ yr., $SD = 10.1$ yr.).

Most of the schizophrenic patients were receiving tranquillizing drugs in the course of treatment. Patients whose medication could not be suspended (about 35%) were eliminated and schizophrenics were randomly allocated to medicated and nonmedicated groups. The medicated groups were maintained on current drugs. For the nonmedicated

groups, drugs were suspended 30 hr. prior to testing. The nonmedicated condition was designed to eliminate some of the gross drug effects but should not be interpreted as showing performance in the complete absence of drug effects.

Control Ss were psychiatric nurses from Gladesville Hospital, New South Wales, Australia (a mean age of 27.4 yr., $SD = 7.2$ yr.).

Apparatus. Eye-movement recordings were made using the EDL Reading Eye camera (Taylor, 1960).

The stimulus display for the free-search condition consisted of two white 35-mm. squares with centers 100 mm. apart on a 14×9 cm. midgray screen and viewing distance of 38 cm.

Size matching was made for two circular patches of light projected onto a 15×10 cm. white screen, the left standard circle was 40 mm. diameter and the right comparison circle was adjusted by S. Distance between the two circles was 10 cm. and viewing distance was 38 cm. The two circles were produced by identical Prado 35-mm. slide projectors with 200-w. white bulbs and circular apertures replacing the slide cartridges 60 cm. behind the translucent screen. White reduction plates with 18 mm. diameter circular holes were placed in front of the final lenses. For the comparison stimulus, an iris diaphragm was used to produce a variable sized smooth circular patch of light. The variable aperture was activated either by a handle on S's right or from the rear by E.

Procedure. After preliminary explanations and establishment of rapport, all Ss were tested on the free-search condition and then on the size-adjustment task under pretraining, training, and extinction conditions.

A practice trial with no specific instructions and a blank stimulus field was given to familiarize Ss with the camera operation. There was a free-search test trial of 20 sec. in which S was instructed simply to look at the display. Previous testing indicated eye-movement measures based on this period were reliable.

After a rest period of 2 min. there were 2 practice trials followed by 6 test trials (pretraining). A further 1-min. rest period was followed by 15 training trials in which E responded to each adjustment with "you are correct" (for the range 38-42 mm. diameter of the comparison circle) or "you set the adjustable circle too small" or "you set the adjustable circle too large." Then there were 6 test trials (posttraining), an additional 15 trials, and a third set of test trials (extinction), all without verbal feedback. All trials were identical size-adjustment trials, except that during the 15 training trials Ss received verbal feedback, and that during the three sets of test trials eye movements were recorded. The starting points for ascending and descending test trials for the comparison circle were 30, 50, 35, 45, 30, 50 mm. in diameter and the intertrial interval was 10 sec.

Eye-movement recordings were made for the free-search and three sets of six size-adjustment test trials and the following measures were obtained:

free search and size adjustment—(a) movement rate: number of large excursion eye movements between the two stimuli per minute; (b) fixation rate: number of fixations or fixation changes per minute; size adjustment only—(c) adjustment rate: number of completed size adjustments per minute (inversely related to mean judgment time). (d) number of movements: number of large eye movements per completed size adjustment. (e) number of fixations: number of fixations per complete size adjustment.

RESULTS

Free-Search Condition

The means for movement rate and fixation rate for the six S groups are shown in Table 1. It appeared that the nonmedicated groups were similar but that the two groups receiving medication showed lower scanning rate.

Since the six S groups were not matched for age, age scores were employed as covariates in the analyses. Factorial analyses of covariance, Six S Groups \times Sex, were made separately for movement and fixation rate measures. Post hoc contrasts for S groups were tested using the R technique (Rodger, 1967) with alpha level .05. The main effect (maximized contrast) for the six S groups was significant for both measures (movements: $F = 1.87$, $df = 5/71$, $p < .05$; fixations: $F = 1.85$, $df = 5/71$, $p < .05$). The eye-movement rates and fixation rates of the two medicated groups were significantly lower than the nonmedicated groups. None of the contrasts between the four nonmedicated groups including the normal control group were significant. The main effect for sex was significant only for the movement-rate scores, with females showing higher movement rate than males (movements: $F = 7.38$, $df = 1/71$, $p < .05$; fixations: $F = .84$, $df = 1/71$, $p > .05$). Interaction effects of S Groups \times Sex were not significant.

These results provide no support for the hypothesis that schizophrenics show extreme scanning behavior in the free-search condition.

Size Judgments

The mean size-judgment scores for the six S groups under pretraining, posttraining, and extinction conditions are shown in Table 2. It appeared that group differences shown for the pretraining condition were eliminated by

TABLE 1
MEANS FOR MOVEMENT AND FIXATION RATE, FREE SEARCH

Measures	Group					
	Nonparanoid schizophrenic medicated	Nonparanoid schizophrenic nonmedicated	Paranoid schizophrenic medicated	Paranoid schizophrenic nonmedicated	Psychotic depressive nonmedicated	Normal control
Movement rate						
Male	21.7	37.7	26.3	29.9	28.5	37.4
Female	30.6	38.2	32.5	46.4	48.1	42.4
Combined	25.6	37.9	29.3	38.1	38.3	39.9
Fixation rate						
Male	60.0	87.5	52.0	81.0	82.5	88.0
Female	72.5	79.5	75.9	90.6	86.5	80.0
Combined	66.2	83.5	63.8	85.8	84.5	84.0

the training procedure. There was a greater spread of size-judgment means for the extinction condition, being intermediate between pretraining and posttraining judgment means.

Because of the extreme changes in variance of sources for different conditions, analyses of covariance were made for each condition separately. The maximized contrast for the six *S* groups was significant only for the pretraining condition ($F = 1.94$, $df = 5/71$, $p < .05$).

Differences in mean size judgments between paranoid and nonparanoid schizophrenic groups (medicated, nonmedicated, and combined) were not significant. For nonmedicated groups, nonparanoid schizophrenics showed greater underestimation than paranoid schizophrenics, whereas for medicated groups, paranoid schizophrenics showed greater underestimation. The contrast for this overall interaction effect was significant. Only the medicated paranoid schizophrenic group showing marked underestimation was significantly different from the normal control group. No significant differences were found between the

mean size judgments of psychotic depressives and schizophrenic patients.

Two additional size-judgment measures were analyzed: the *absolute-error* measure (absolute value of the constant error) and the *intra-S variance* of size adjustments. Means for the six *S* groups under training conditions are shown in Table 3. For both measures the maximized contrasts for the six *S* groups were significant for pretraining, posttraining, and extinction conditions (errors: pretraining $F = 4.42$, $df = 5/71$, $p < .05$; posttraining $F = 2.59$, $df = 5/71$, $p < .05$; extinction $F = 3.51$, $df = 5/71$, $p < .05$; variance: pretraining $F = 1.71$, $df = 5/71$, $p < .05$; posttraining $F = 2.09$, $df = 5/71$, $p < .05$; extinction $F = 2.09$, $df = 5/71$, $p < .05$). Except in two instances all patient groups showed significantly greater absolute errors and intra-*S* variance than the control group for all conditions and there were no significant differences between patient groups.

Despite the evidence that the training procedure eliminated differences in mean size

TABLE 2
MEANS OF SIZE JUDGMENTS FOR PRETRAINING, POSTTRAINING, AND EXTINCTION

Condition	Group					
	Nonparanoid schizophrenic medicated	Nonparanoid schizophrenic nonmedicated	Paranoid schizophrenic medicated	Paranoid schizophrenic nonmedicated	Psychotic depressive nonmedicated	Normal control
Pretraining	-1.21	-2.60	-4.10	.38	-2.07	-.69
Posttraining	-.64	-.60	-.57	-.42	-1.00	-.37
Extinction	-1.46	-2.60	-.62	-.40	-1.46	-.40

Note.—Constant error in millimeters.

TABLE 3
MEANS OF ABSOLUTE ERROR AND INTRASUBJECT VARIANCE SCORES

Measure and condition	Group					
	Nonparanoid schizophrenic medicated	Nonparanoid schizophrenic nonmedicated	Paranoid schizophrenic medicated	Paranoid schizophrenic nonmedicated	Psychotic depressive nonmedicated	Normal control
Absolute error						
Pretraining	9.06	10.37	14.77	10.49	9.17	2.21
Posttraining	4.06	3.21	2.86	3.01	5.03	1.68
Extinction	8.63	10.11	5.58	7.06	9.26	1.72
Intrasubject variance						
Pretraining	51.1	36.2	47.5	43.5	30.4	17.2
Posttraining	46.7	27.9	16.5	26.6	30.6	12.8
Extinction	32.2	32.7	27.6	47.1	33.8	12.9

judgments for the six S groups, the greater absolute errors and intra-S variance of size judgments of patients compared with control Ss were not eliminated.

Eye-Movement Scores

Results for the five eye-movement measures were treated separately in analyses of covariance for the pretraining, posttraining, and extinction conditions. The *F* ratios for the maximized contrasts for means of the six S groups are shown in Table 4. The group means for the scores with significant maximized contrasts are shown in Table 5.

None of the 15 contrasts testing the differences between paranoid and nonparanoid groups (medicated, nonmedicated, and combined) was significant. There was no evidence that schizophrenic patients showed extreme scanning defined either by rate of eye movement (Measures 1 and 2) or number of movements (Measures 3 to 5).

There were significant differences between patients and the control group for movement-

rate scores under posttraining and extinction conditions. However, the lower movement rate for patient groups as a whole was mainly due to the two medicated groups supporting the results for the free-search condition. For the four nonmedicated groups there were no consistent differences between paranoid and nonparanoid schizophrenics and the normal control group or the psychiatric control group of psychotic depressives on any of the eye-movement scores.

Relation between Size Judgments and Eye-Movement Scores

It could be argued that failure to confirm differences between paranoid and nonparanoid schizophrenics and normals was partly a function of difficulties of psychiatric classification for patient groups. If all schizophrenics are examined regardless of subtype, then from scanning theory a relation is predicted between extreme scanning behaviors and size overestimation and underestimation. It was decided to examine the correlations of eye-movement and size-judgment scores for the whole patient group. Obviously the methods of selection of the patient groups and combining medicated and nonmedicated patients restrict interpretation of the correlations. The intercorrelations of the two sets of scores for the pretraining condition are shown in Table 6.

The pattern of eye-movement and size-judgment correlations was similar to those found in a more extensive study with normal Ss (McKinnon, 1966). None of the 45 corre-

TABLE 4
F RATIOS FOR MAXIMIZED CONTRASTS
OF SUBJECT GROUP MEANS

Measure	Pre- training	Post- training	Extinction
Movement rate	1.05	1.74*	2.45*
Fixation rate	1.16	1.26	.82
Adjustment rate	.56	1.52	.60
Amount of movement	2.04*	.74	1.33
Amount of fixation	2.27*	.58	.71

Note.— $F(1 - \alpha)^2$ (Rodger, 1965). $F = 1.43$, $df = 5/71$.

* $p = .05$.

TABLE 5
GROUP MEANS FOR SCORES WITH SIGNIFICANT CONTRASTS

Measure and condition	Nonparanoid schizophrenic medicated	Nonparanoid schizophrenic nonmedicated	Paranoid schizophrenic medicated	Paranoid schizophrenic nonmedicated	Psychotic depressive nonmedicated	Normal control
Movement rate posttraining	44.4	58.6	48.8	58.4	53.5	76.8
Movement rate extinction	48.5	67.5	50.2	65.4	50.9	79.5
Adjustment rate posttraining	10.1	9.9	8.5	11.5	9.6	12.4
Number of movements pretraining	6.4	11.5	7.4	9.3	7.7	7.7
Number of fixations pretraining	11.2	19.0	12.5	14.9	15.2	11.9

lations of eye-movement and size-judgment scores was significant at the .05 level.

DISCUSSION

The results failed to support the experimental hypotheses predicting differences in size-judgments and visual scanning behavior of paranoid and nonparanoid schizophrenics and psychotic depressives and control Ss.

Findings for mean size judgments of medicated schizophrenics were consistent with results reported by Silverman (1964b); medicated paranoid schizophrenics showed greater underestimation than medicated nonparanoid schizophrenics. However, the opposite results were found for nonmedicated groups. From information available in the present experiment it is not possible to interpret these effects. Although most of the schizophrenic patients were receiving phenothiazine drugs (as in the Silverman study), there may have been systematic differences between paranoid and nonparanoid groups in the particular medication and dosage parameters.

Although the procedure used to assess the effects of medication involved limited control, there were differences between medicated and nonmedicated patients for both size-judgment and visual scanning scores. Tranquillizer medication produced a reduction in visual scanning rate for paranoid and nonparanoid schizophrenic patients. The results for drug effects indicated the possibility of confounding drug effects in several previous studies on schizophrenics with tasks taken as indicating scanning behavior.

Errors shown by psychotic patients in size-judgment mean scores were eliminated by the training procedure involving verbal reinforcement. However the greater variability shown by patients, variability for each individual from occasion to occasion, and variability between individuals was not eliminated by the training procedure.

Schizophrenics and psychotic depressive patients performed less adequately than control Ss on the size-adjustment task but the deficit was not systematically related to differences

TABLE 6
CORRELATIONS OF SIZE JUDGMENTS AND EYE MOVEMENT SCORES

Item	2	3	4	5	6	7	8
Size judgment mean	-.60	-.13	.11	.05	.17	-.14	-.18
Nondirection error		-.06	.02	.00	-.30	.27	.31
Intrasubject			-.16	-.24	.03	-.09	-.08
Movement rate				.58	.03	.49	.16
Fixation rate					.27	.13	.16
Adjustment rate						-.70	-.74
Number of movements							.88
Number of fixations							

Note.—N = 70.

in direct eye-movement measures of visual scanning. When medication effects were controlled there was no evidence that schizophrenic patients showed extreme scanning dispositions defined by direct measures of visual scanning. Finally, because of the stability and generality of individual differences previously demonstrated for visual scanning measures the conclusions can be applied to a wide range of situations involving visual scanning.

REFERENCES

- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. L., & SPENCE, D. W. A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1(4), 1-186.
- GARDNER, R. W., & LONG, R. I. Cognitive controls of attention and inhibition: A study of individual consistencies. *British Journal of Psychology*, 1962, 53, 381-388. (a)
- GARDNER, R. W., & LONG, R. I. Control, defence and the centration effect: A study of scanning behavior. *British Journal of Psychology*, 1962, 53, 129-141. (b)
- MCKINNON, T. A. Individual differences in visual scanning behavior. Unpublished master's thesis, University of Sydney, 1966.
- RODGER, R. S. *Intermediate Statistics*. Sydney: University Bookshop, 1965.
- RODGER, R. S. Type I Errors and their decision basis. *British Journal of Mathematical and Statistical Psychology*, 1967, 20, 51-62.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 352-380. (a)
- SILVERMAN, J. The scanning control mechanism and 'cognitive filtering' in paranoid and non-paranoid schizophrenics. *Journal of Consulting Psychology*, 1964, 28, 385-393. (b)
- STANNARD, R., SINGER, G., & OVER, R. The effect of operant conditioning on size judgments of schizophrenics. *British Journal of Psychology*, 1966, 57, 329-334.
- TAYLOR, S. E. *Eye movement photography with the reading eye*. New York: Educational Development Laboratory, 1960.

(Received November 6, 1967)

SOCIOMETRIC STUDY OF THE SEX OFFENDER

MURRAY COHEN¹

Boston University

THEOHARIS SEGHNOR

Treatment Center, Bridgewater, Massachusetts

AND WILFRED CALMAS

Boston State College

A most frequent clinical description of the sexual offender is his deficiency in social skills. In studying these defects, however, the use of broad medicolegal groups as a parameter is a questionable procedure. In this exploratory study 65 pedophiles and rapists, living in a segregated treatment center, were identified by a diagnostic procedure and placed into a clinical descriptive classification. A sociometric procedure was used to study differences in social effectiveness and to test the usefulness of the classification. The diagnostic procedure identified four groups of rapists and three groups of pedophiles. The results indicate that the clinical classification is able to clarify sociometric differences and is of greater research value than the medicolegal grouping.

From a clinical frame of reference an obvious feature of the sexually deviant act is its essentially asocial character. The sexual act in the normal adult involves a mutuality of experience, a reciprocal sensitivity, and concern of the one for the other. By contrast, the acts of rape, child assault, exhibitionism, voyeurism, fetishism are basically impersonal and egocentric.

In the course of studying over 800 male, young adults and adults who had committed sexually deviant acts, it became apparent that among other disturbances there is also present a marked defect in interpersonal relationships with a relative absence of even the most basic of social attitudes and values or social skills. Such a finding has been reported in other clinical and research studies (Allen, 1962; Gebhard, Gagnon, Pomeroy, & Christenson, 1965; Glueck, 1956; Henry, 1941; Karpman, 1954).

The study of these patients occurred as a result of a law passed by the Massachusetts General Court (General Laws, 1958) which established a center for the diagnosis and treatment of convicted sexual offenders (Cohen & Kozol, 1966; Kozol, Cohen, & Garofalo, 1966). Under this law a person found guilty of a sexual offense can be committed through

a civil procedure from one day to life in lieu of receiving a criminal sentence. Predatory homosexuals with some age disparity between the offender and victim and father-daughter or brother-sister incest cases have been committed under the law. The majority of the committed patients, however, are pedophiles and assaultive rapists.

Although all show some defect in relating to others, clinical observations point to differences in the nature, level, and quality of interpersonal experiences and social skills. In the Sing Sing study (Glueck, 1956), the authors concluded that rapists were more capable in social interactions than pedophiles and went on to describe the latter group as immature and socially inadequate. Similar statements about the pedophile can be found in other clinical studies of sexual offenders (Allen, 1962; Bromberg, 1948; Gebhard et al., 1965; Guttmacher, 1951; Karpman, 1954; Mohr, Turner, & Jerry, 1964). The authors' work, however, led them to believe that the use of a medicolegal classification of sex offenders as a parameter in research is clinically and methodologically unsound. The purpose of this paper is to present a brief description of a clinically derived classification of sexual offenders and to report the results of a sociometric study designed to test some implications of the classification for its clinical, theoretical, and research utility.

¹ Request for reprints should be sent to Murray Cohen, Department of Psychology, Boston University, 111 Cummington Street, Boston, Massachusetts 02215.

Clinical Classification

Rape. Descriptively, the act of rape involves both sexual and aggressive features. It is clear, however, that there are different relationships between these two motives expressed in the sexual assault.

In some acts of rape the intent of the act is primarily aggressive with sexual feelings minimal or absent. The sexual behavior is used to physically harm, to degrade, or to defile the victim in the service of this aggressive intent. The acts are experienced by the offender as the result of an "uncontrollable impulse" and almost always follows some precipitating, disagreeable event involving a wife, girlfriend, or mother. The victim is brutally assaulted, and those parts of the woman's body which usually are sexually exciting frequently become the foci for the offender's violence.

Sexual excitation itself is often absent or only minimally present. The offender must masturbate to get an erection and in a large number of cases, he cannot reach an orgasm. It is readily apparent that the sexual behavior is not meant to gratify an intense sexual desire but rather is serving an aggressive motive. Since, in all such cases, the victim is unknown to the offender, and therefore could not herself be the incentive for the anger, she is apparently the object for displaced aggression. Such patients therefore are classified as rapist-displaced-aggression type.

In other acts of rape the aggressive features are less violent and frequently play such an inappreciable part, that should the victim struggle with any degree of vigor, the offender flees. In these acts the primary aim is clearly sexual and the aggression is in the service of gratifying the sexual desires. The offender is always in a state of intense sexual excitation and often has an orgasm in the simple pursuit of the victim or upon making some physical contact. The recurrent fantasy in such offenders is that the victims will yield, submit to intercourse, in which he will be especially virile and so pleasing to the victim that she will become enamoured with him and invite him to repeat the sexual acts. Outside of these episodic sexual assaults, there is no other antisocial behavior and in fact such

offenders are extremely passive and submissive. There is a pervasive, almost obsessive, concern with feelings of sexual inadequacy and the assault appears to be an effort to compensate for or to counteract these feelings of impotency and inadequacy. These patients are classified as rapist-compensatory type.

In a third group, the sexual and aggressive desires appear to coexist so that the offender is not able to experience, or even fantasy, sexual desires without a concomitant arousal of aggressive thoughts and feelings. Further, he projects such feelings onto his victim and sees her struggles and protestations as seductive—"Women like to get roughed up; they enjoy a good fight." Such perceptions are made in the context of very brutal assaults. In offenders where this pattern is present, all object relationships, with both men and women, show this same quality of an eroticization of aggressive behavior. Such patients are very loud and assertive and those who come in contact with them feel both overwhelmed and seduced at the same time. The extreme of this position is sexual sadism where the woman is viciously violated and murdered. Because the aggression in the sexual act is so primitive, lacking any neutralization by the sexual feelings, this offender is classified as rapist-sex-aggression-defusion type.²

In a final group of rapists neither sexual nor aggressive desires appear to play a significant role. Rather, the act appears as simply another aspect of the offender's predatory nature. The rape is frequently carried out in the context of some other antisocial act such as robbery or theft. In such instances the victim is unfortunately present and is sexually assaulted with relatively little sexual or aggressive feelings. The act is opportunistic, narcissistic, and impulsive, and such an offender has a history of predatory, antisocial behavior from his late childhood, preadolescent years. This group of sexual offenders are classified as rapist-impulse type.

Pedophilia. In pedophilia the sexual object

² For a discussion of sex-aggression defusion see Nunberg (1955, pp. 102-113). He does not address himself to sexual deviancy in this discussion, but deals with the concept in terms of the psychoanalytic theory of instinctual development.

is the prepubertal child. In the authors' clinical studies three groups of such offenders were differentiated on the bases of descriptive features of the act, the intentionality of the act and the developmental history of the offender.

In one group of child offenders, the sexual interests are expressed as desires to touch, fondle, caress, suck, and smell the child. In this group the offender has never been able to develop or maintain mature object relationships with his male or female peers at any stage of his adolescent, young adult, or adult life. He is socially comfortable only with children and seeks them out as companions or seeks out employment where he can work or be near them (camp counselors, church work, YMCA). In most instances the child is known to the offender. There is nothing impulsive about the act for he begins the relationship by a type of disarming courtship which engages the child and often the parents of the child. The sexual play occurs only after a period of seduction. His social behavior is best described as passive dependent. Because of the apparent arrest in psychosocial development such patients have been classified as pedophile-fixated type.

In another group of child offenders the sexual aim is also primarily pregenital although here, unlike the fixated pedophile, there is a history of an apparently normal adolescence with good peer relationships, some dating behavior, and heterosexual experiences. However, throughout this period there exist increasing feelings of masculine inadequacy in sexual and nonsexual activities. And, as he enters adulthood, his social, occupational, and marital adjustment is quite tenuous and marginal. There is frequently a history of an inability to deal with the normal stresses of adult life and alcoholic episodes become increasingly more frequent and result in the breakdown of a relatively stable marital, social, and work adjustment. In almost all instances the pedophilic acts are precipitated by some direct confrontation of his sexual adequacy by an adult female or some threat to his masculine image by a male peer. Most frequently, however, the precipitating event is the discovery that his wife or girlfriend is having an affair with another man. Although

a variety of different dynamics are present in the different offenders within this group, a similar consequence to the sexual rejection occurs which can best be described as an aim and an object regression. Such patients, therefore, are classified as pedophile-regressed type.

There are two additional differences between offenders in this group and offenders of the fixated type. In those instances where regression is seen to take place, the victim is not a familiar neighborhood child and the act is characteristically impulsive. The offender will describe driving or walking along the street, seeing a child and suddenly becoming overwhelmed by sexual excitation which he cannot control. Second, in almost all cases, the victim of the offenders in the pedophile, regressed group is a female child, with homosexual pedophilia extremely rare.

In a third group of child offenders the aim of the relationship with the child contains both sexual and aggressive features. Although this is superficially analogous to the sex-aggression-defusion rapist, the object and the dynamics are quite different. The majority of offenders in this group select boys as the object for the sexual sadism, and the sexual component is far less object oriented. The primary aim is aggression, and is expressed in cruel and vicious assaults on the genitalia or by introducing the penis or elongated objects into the victim orally or anally. The sexual excitement increases as an apparent function of the aggression, but the orgasm itself either does not occur or must be reached through masturbation. These patients are classified as pedophile-aggressive type.

The clinical classification was developed empirically and is therefore primarily descriptive and atheoretical. However, the classification appears to yield expectations of differences among the subgroups with respect to social adaptiveness and social skills; differences which are obscured by the broader medicolegal grouping.

The general hypothesis of the present study is that an analysis of interpersonal social skills in terms of the clinical subclasses is more sensitive than an analysis based upon the medicolegal classification.

Further, two specific predictions are sug-

gested by the classification. As the rapist-displaced-aggression type and the pedophile-regressed type have demonstrated developmentally higher levels of social adaptation, and the sexual acts appear to be reactive and are experienced as dystonic, it would be expected that they would show the highest levels of social skills among the sexual offenders—more popular, more socially active, and more socially sensitive. Second, since the pedophile-fixated type, the pedophile-aggressive type, and the rapist-impulse type appear to be fixated at early levels of object relationships, and the sexual acts represent characteristic ways of dealing with the social world, it would be expected that they would show few social skills and have little social success.

METHOD

The Treatment Center, Bridgewater, Massachusetts, is a segregated unit within a correctional institution and is a relatively closed social system. By virtue of this, a sociometric procedure appeared ideally suited for a study of these derivations. At the time of the study, there were 84 committed patients at the Center. The data for the research were obtained from 72 patients, an 85% sample, with the remaining 12 patients refusing to participate. Seven patients were committed for incestuous acts with daughters or sisters and although their choices were included as part of the sociometric data for the scores of the other groups, this incest group was not included in the analyses. Of the remaining 65 patients, 27 were committed for acts of rape: 10 were classified as rapist-displaced aggression, 4 rapist-compensatory, 4 rapist-sex aggression defusion, and 9 rapist-impulse. Thirty-eight were committed for sexual acts against children: 23 were classified as pedophile-fixated, 8 pedophile-regressed, and 7 pedophile-aggressive.

All patients had been together for a period ranging from 6 mo. to 5 yr. The patients were seen individually and given a sociometric questionnaire. They were asked to give the names of three patients whom they would prefer to be with in each of four activities and three patients whom they would not want to associate with in these same activities. The activities included group therapy, leisure time, work, and the dining room. In addition, they were asked to select three patients whom they preferred for patient government roles and three whom they would not want to have in such roles.

For each patient seven sociometric measures were derived. Although these measures are not completely independent of each other, each variable was felt to contribute some information not contained in the other variables. The measures include (a) positive choice status—the number of positive choices received by each patient; (b) negative choice status—the number of negative choices received by each

patient; (c) net choice status—the algebraic sum of the first two variables; (d) total activity—the arithmetic sum of positive and negative choice; (e) asymmetrical reciprocity—the number of times a patient selected another patient who either omitted him from a positive choice or selected him as a negative choice; (f) symmetrical reciprocity—the number of times that a patient was chosen by another patient that he himself had chosen; and (g) index of attraction—a ratio score derived from the number of positive choices divided by total activity.

On preliminary analysis of the data it was found that all five choice situations, (group therapy, leisure time, dining room, work, and patient government) showed significantly high correlations with each other. Such findings are consonant with the results of studies which point to the fact that one individual tends to see another as a whole and either likes or dislikes him resulting in a uniform choice pattern. The finding of this "halo" effect which influences choice on all activities led to a collapsing of activity categories and a consideration of the sociometric variables independent of the particular social activity.³

The data were statistically evaluated by *t* tests for differences between the two medicolegal groups and analyses of variances for differences among the various subgroups. Where, in the analyses of variance, significant differences were found, a multiple-range test developed by Duncan (1955) and extended for use with groups of unequal size by Kramer (1956) and Duncan (1957) was applied to test for significance of differences among the subgroup means. This most conservative test was applied because of a number of methodological weaknesses in the research, which despite this corrective effort, must make for caution in generalization. The most important defects include the problem of sample representativeness, the small number of observations in some of the subgroups, the empirical heterogeneity of patients within subgroups, and the inability to make specific hypotheses for each variable for each subgroup.

RESULTS

The data for all variables for each of the clinical groups, as well as for the overall groups of rapists and pedophiles, are presented in Table 1. For the sake of convenience, Table 2, which is derived from Table 1, presents a ranking of mean scores of the clinical groups, ordered on the basis of a rank score of 1 representing the highest level of social adaptation. The final column of Table 2 represents the overall ranking of the clinical groups across all variables. Since it is not

³ In a replication study now being carried out, separate analyses of the activities will be performed to determine what, if any, contribution such analyses will make.

TABLE 1
MEANS FOR THE SOCIOMETRIC VARIABLES

Clinical subgroup ^a	N	Sociometric variables													
		Positive status		Negative status		Net status		Total activity		Asymmetrical reciprocity		Symmetrical reciprocity		Index of attraction	
		\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
Pedophile															
PF	23	9.35	7.33	13.12	16.75	-3.77	20.58	22.47	16.55	12.30	2.25	2.96	2.18	41.61	31.55
PR	8	12.62	8.35	5.38	5.08	7.24	11.05	18.00	9.79	11.00	1.92	4.14	1.73	70.11	26.78
PA	7	5.15	7.93	15.71	12.09	-10.56	14.57	20.86	10.06	13.29	1.80	1.86	1.77	24.69	21.30
Combined	38	9.89	7.38	12.60	14.46	-2.71	17.82	21.23	14.10	12.21	2.18	3.00	2.10	46.62	30.76
Rapist															
RD	10	25.20	12.37	5.70	4.70	19.50	15.09	30.90	12.52	9.30	4.59	5.00	3.35	81.55	15.39
RC	4	8.75	5.94	4.00	4.25	4.75	7.68	12.75	2.05	13.75	1.24	2.00	1.32	68.63	31.28
RSA	4	6.00	3.74	8.33	8.39	-2.33	8.96	14.33	5.86	14.00	.00	1.66	.80	41.87	32.84
RI	9	9.22	8.40	19.11	17.37	-9.89	22.82	28.33	15.56	11.78	3.07	3.44	1.74	32.55	32.65
Combined	27	13.33	12.56	9.05	12.47	4.28	20.45	24.88	13.88	11.38	3.79	3.62	2.63	53.60	31.64

^a Pedophiles: PF = fixated; PR = regressed; PA = aggressive; Rapists: RD = displacement; RC = compensatory; RSA = sex-aggression defusion; RI = impulse.

clear that higher activity is necessarily associated with better social adaptation, Table 2 contains no ranking for this variable.

The data were first analyzed to evaluate differences between the medicolegal grouping of rape and pedophilia, on each variable. For this analysis a series of *t* tests was performed. The results, presented in Table 3, show that no variable discriminated between the two medicolegal groups. That there is no clear differentiation between the medicolegal groups can also be noted by referring to Table 2. It can be seen that certain subgroups of rapists and pedophiles rank higher or lower than subgroups of the other general offense category.

TABLE 2
RANK ORDERING OF SUBGROUPS ON THE BASIS OF SOCIAL ADAPTATION

Clinical subgroup	Sociometric variables						
	Positive status ^a	Negative status	Net status	Asymmetrical reciprocity	Symmetrical reciprocity	Index of attraction	Rank of sum of ranks
PF	3	5	5	4	4	5	4
PR	2	2	2	2	2	2	2
PA	7	6	7	5	6	7	7
RD	1	3	1	1	1	1	1
RC	5	1	3	6	5	3	3
RSA	6	4	4	7	7	4	6
RI	4	7	6	3	3	6	5

^a Rank 1 indicates the higher level of social adaptation.

A series of analyses of variance was then carried out to determine if the obtained mean differences of the subgroup samples were statistically significant. A summary of the analyses for each sociometric variable appears in Table 4. For five of the seven variables (positive choice, net choice, asymmetrical and symmetrical reciprocity, and index of attraction) the differences between the clinical subgroups are significant at or beyond the .05 level of probability. Only negative choice and total activity failed to discriminate statistically between the groups.

In order to clarify further the obtained differences the multiple-range test was applied to the subgroup data and the results of this analysis are shown in Table 5.

With respect to positive choice status, it can be seen that the rapist-displaced-aggression group is significantly differentiated

TABLE 3
SIGNIFICANCE OF DIFFERENCES BETWEEN MEDICOLEGAL GROUPS

Sociometric variable	Rapists		Pedophiles		<i>t</i> ^a
	\bar{X}	SD	\bar{X}	SD	
Positive status	13.33	12.56	9.89	7.38	.37*
Negative status	9.05	12.47	12.60	14.46	1.40***
Net status	4.28	20.45	-2.71	17.82	1.45***
Total activity	24.88	13.88	21.23	14.10	1.41***
Asymmetrical reciprocity	11.38	3.79	12.21	2.18	1.11**
Symmetrical reciprocity	3.62	2.63	3.00	2.10	1.10**
Index of attraction	53.60	31.64	46.62	30.76	1.06**

^a *df* = 63.

* *p* = .40.

** *p* = .20.

*** *p* = .10.

TABLE 4

ANALYSES OF VARIANCE FOR THE VARIABLES

Sociometric variable	Mean square between groups ^a	Mean square within groups ^b	F	p
Positive status	457.7	66.1	6.92	.005
Negative status	266.1	179.7	1.48	.25
Net status	1052.2	317.0	3.31	.01
Total activity	610.4	281.7	2.16	.10
Asymmetrical reciprocity	19.7	7.5	2.63	.025
Symmetrical reciprocity	11.1	4.8	2.29	.05
Index of attraction	2824.5	809.4	3.49	.005

^a df = 6.
^b df = 58.

from all other subgroups on the basis of a higher positive score. The remaining subgroups are statistical equivalent on this variable. For net choice status, the rapist-displaced-aggression subgroup is significantly differentiated from the rape-impulse, pedophile-aggressive, and pedophile-fixated subgroups. And again all other subgroups are undifferentiated. On both the asymmetrical and symmetrical reciprocity variables the rapist-displaced-aggression group differs significantly from both the rapist-sex-aggression-defusion and rapist-compensatory groups. Furthermore, for asymmetrical reciprocity the rapist-displaced-aggression group also shows a significantly higher score than the pedophile-aggressive group. Finally, on the index of attraction dimension, two significant discriminations obtained. First, the rapist-displaced-aggression group has a higher index of attraction than all other clinical subgroups, and, second, the rapist-compensatory group scores significantly higher than the pedophile-aggressive group.

DISCUSSION

The findings support the general hypothesis of the study. The analysis of sociometric behavior in terms of clinical subclasses is more sensitive than such an analysis applied to broad medicolegal classes. Further, the findings strongly support the first specific prediction and to a lesser degree the second. The rapist-displaced-aggression group is clearly functioning at the highest level of social effectiveness as compared with all other sexual offenders. And, also in accord with the first prediction, the pedophile-regressed group follows closely behind on all the sociometric scales.

With respect to the second prediction, the pedophile-aggressive group consistently shows ineffective social functioning as expected, but the findings for the rapist-impulse group and the pedophile-fixated group are more equivocal. The rapist-impulse group is comprised of offenders who have a long history of a variety of antisocial, predatory acts. Clinically, they are described as impulse ridden and self-centered with no ability to form meaningful object relationships and frequently diagnosed in classical nosology as psychopaths or sociopaths. And yet, according to the present findings, they have some social skills and sensitivity to an unexpected degree. Similarly, the the pedophile-fixated group showed greater social adaptability than was expected. Patients are placed in this group when there is a history of a relative absence of mature object relations with either sex. Clinically, they appear as passive-dependent persons, socially inadequate with intense feelings of isolation and alienation. Despite this picture they are not that socially isolated and some social skills are obviously present.

TABLE 5
MULTIPLE COMPARISONS OF SUBGROUP MEANS

Sociometric variable	Multiple comparisons ^a
Positive status	(RD) (PA, RC, RI, PF, PR, RSA)
Negative status	No significant differences
Net status	(RD, RC, PR) (RC, PR, PA, RI, PF, RSA)
Total activity	No significant differences
Asymmetrical reciprocity	(RD, RI, PR, PF, PA) (RSA, RC, RI, PR, PF, PA)
Symmetrical reciprocity	(RD, RI, PF, PR) (RSA, PA, RC, RI, PF, PR)
Index of attraction	(RD, RC, PR) (PA, RI, RSA, PF, PR) (RC, RI, RSA, PR)

^a * Any two groups not appearing together within the same parentheses have significantly different means ($p \leq .05$). Any two groups appearing together within the same parentheses have means which do not differ ($p > .05$) (after Duncan, 1956).

The rapist-compensatory group shows the greatest amount of variability of social effectiveness. Although relatively socially insensitive (a low rank on both asymmetrical and symmetrical reciprocity) and most socially isolated (lowest total choice), such patients are least offensive or least actively disliked (least chosen as negative choice). This finding supports the clinical observations. Rape, as a compensatory act, is seen to take place in men who have a sense of impotence regarding their impact on the world. Such an ego attitude is expressed in the passive and submissive manner in which they relate to others. This mode of relating helps them avoid direct confrontations but makes them, in the social world, powerless. Thus, they are inoffensive, but the cost is social isolation.

It is clear that the clinically observed heterogeneity of sexual offenders can be tested by behavioral research. This is not, as it appears to be, stating the obvious, for with some rare exceptions, despite the voluminous literature on normal and deviant human sexual behavior, the published material is made up of clinical anecdotal reports, broad ranging statistical descriptive studies, narrow theoretical exegeses, or exhortative tracts of a moral or religious nature. Though there have been some well-executed cultural anthropological studies (Malinowski, 1929; Mead, 1949) and a number of systematic physiological hormonal studies (Beach, 1948; Ford & Beach, 1951), behavioral research has been lacking to a notable degree.

One final point should be made. Clinical classification has practical as well as theoretical and research implications. A primary applied application of valid classification lies in treatment planning and prognosis. It is quite clear, for example, even from the brief discussion of the characteristics of the groups, that the rapist-displaced-aggression and the rapist-impulse groups cannot be treated in the same way, and that prognosis is quite different. For the latter group the sexual problem is secondary to the general inability to adjust to social demands and to effect a close, positive, interpersonal relationship. For the patients so diagnosed, formal psychotherapy must take place in a context of an overall effort at socialization. And, as the antisocial behavior has had such an early onset (pre-

pubertal years) and pervades so many aspects of the patient's life, the treatment will be a long and arduous process. In those acts of rape, however, where the major dynamic appears to be a focalized displaced aggression, where a therapeutic alliance is relatively easy to establish, and where the patient has a large number of social and adaptive skills, classical therapeutic techniques can be applied with positive results expected within a relatively short period.

REFERENCES

- ALLEN, C. A. *A textbook of psychosexual disorders*. London: Oxford University Press, 1962.
- BEACH, F. A. *Hormones and behavior*. New York: Hoeber, 1948.
- BROMBERG, W. *Crime and the mind*. New York: Lippincott, 1948.
- COHEN, M. L., & KOZOL, H. L. Evaluation for parole at a sex offender treatment center. *Federal Probation*, 1966, 30, 50-55.
- DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, 11, 1-2.
- DUNCAN, D. B. Multiple range tests for correlated and heteroscedastic means. *Biometrics*, 1957, 13, 164-176.
- FORD, C. S., & BEACH, F. A. *Patterns of sexual behavior*. New York: Harper, 1951.
- GEHARD, P. H. GAGNON, J. H., POMEROY, W. P., & CHRISTENSON, C. V. *Sex offenders*. New York: Harper, 1965.
- General Laws, Commonwealth of Massachusetts*. Chapter 123A, 1958.
- GLUECK, B. C. Final Report. Research project for the study and treatment of persons convicted of crimes involving sexual aberrations. New York State, Department of Mental Hygiene, 1956.
- GUTTMACHER, M. S. *Sex offenses: The problem, causes, and prevention*. New York: W. W. Norton, 1951.
- HENRY, G. W. *Sex variants*. (2 vols.) New York: Hoeber, 1941.
- KARPMAN, B. *The sexual offender and his offenses*. New York: Julian Press, 1954.
- KOZOL, H. L., COHEN, M. L., & GAROFALO, R. F. The criminally dangerous sex offender. *New England Journal of Medicine*, 1966, 275, 79-84.
- KRAMER, C. V. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, 12, 307-310.
- MALINOWSKI, B. *The sexual life of savages*. London: Routledge, 1929.
- MEAD, M. *Male and female*. New York: Morrow, 1949.
- MOHR, J. W., TURNER, R. E., & JERRY, M. B. *Pedophilia and exhibitionism*. Toronto: University of Toronto Press, 1964.
- NUNBERG, H. *Principles of psychoanalysis*. New York: International Universities Press, 1955.

(Received November 10, 1967)

PUNISHMENT OR DISTRACTION? OPERANT STUTTERING REVISITED

BARBARA BIGGS AND JOSEPH SHEEHAN¹

University of California, Los Angeles

In a partial replication of a study by Flanagan, Goldiamond, and Azrin, six male stutterers enrolled in the Psychology Speech Clinic of the University of California, Los Angeles, were tested during successive readings of a given passage. Three experimental conditions were employed using a 4,000-cps tone at an intensity of 108 db., ISO hearing level, to determine the effect of this stimulus upon stuttering behavior. The conditions were (a) an aversive condition in which the presentation of the tone was contingent upon stuttering; (b) an escape condition in which cessation of the tone was contingent upon stuttering; and (c) a random condition in which the tone was presented independent of Ss' disfluencies. These findings, differing from those of Flanagan et al., were that stuttering decreased in all conditions. The results are attributed primarily to distraction effects.

In 1958, Flanagan, Goldiamond, and Azrin conducted an experiment in "operant stuttering" with three Ss. They employed a 1-sec. blast of a 6,000-cps tone at an intensity level of 105 db.² (Flanagan, 1966) as a noxious stimulus. They found that stuttering increased in frequency when a 5-sec. cessation of a continual tone was made contingent upon stuttering. It decreased when presentation of the stimulus was contingent upon stuttering.

This result is somewhat at variance with those of other studies reported in the literature concerning both the theoretical and actual effects of punishment and reward on verbal nonfluency. These studies detail varied individual reactions depending upon both the strength and type of punishment employed and the characteristics of Ss themselves (Frick, 1951; Hill, 1954; Johnson, 1958; Shames & Sherrick, 1963; Sheehan, 1951, 1953, 1958; Sheehan, Cortese, & Hadley, 1962; Sheehan & Voas, 1957; Siegel & Martin, 1965a, 1965b; Skinner, 1957; Stassi, 1961; Van Riper, 1937).

Since this widely quoted study employed only a few Ss, and since its authors did not consider the possibility that their results

might be due to distraction effects (cited by Bloodstein—1950—as one of the two conditions responsible for almost "all of the reductions in stuttering recorded in this study"), the present study was undertaken as a partial replication, with an increase in the number of Ss and certain modifications in design.

METHOD

Design

Three design changes were undertaken. First there were questions as to the reliability of the stimulus employed by Flanagan, Goldiamond, and Azrin. Since it is common for males with otherwise normal hearing to have some loss of acuity at high frequencies, there was concern about the possible necessity of eliminating potential Ss if losses were found at 6,000 cps. Informal judgments made by E and others also led to questions about the aversive properties of the tone. For these reasons, a pretest was run on normally fluent speakers. As a result, the stimulus selected for the present experiment was a 4,000-cps tone presented at an intensity level of 108 db., ISO³ hearing level.

Second, it has been well established that the specific material chosen for reading has a decided effect on stuttering behavior (Sheehan, 1958; Taylor, 1966). Therefore Ss in the present study were asked to read the same passage throughout a given experimental session, rather than different passages which could cause fluctuations in performance that would be difficult to analyze.

The experimental design was modified by adding a third condition—noncontingent, random presentation of the tone. This specifically deals with the

³ International Standards Organization, 1964, specifications for pure tone audiometers.

¹ Requests for reprints should be sent to Joseph Sheehan, Department of Psychology, University of California, Los Angeles 24, California.

² Audiometric zero as described by American Standards Association, 1951, Specifications for pure tone audiometers (B. Flanagan, personal communication, 1966).

distraction variable; the tone does not "reward" the stutterer by stopping at each disfluency; neither does it appear each time he stutters to "punish" him.

Statistically, an incomplete balanced block design was employed. Order effects were controlled by the random assignment of *Ss* to a schedule of experimental sessions involving every order and sequence of the three conditions investigated.

Subjects

The *Ss* were six male stutterers currently enrolled in the Psychology Speech Clinic at the University of California, Los Angeles. Their ages ranged from 18 to 28 yr., an average age of 24. Participation in the study was voluntary.

Apparatus

Ambient noise in the experimental room was monitored with a General Radio Model 1551-B sound level meter. The sound stimulus was produced by an Allison 22 audiometer, calibrated with an Allison Model 3A Calibration Unit. The *Ss* received the tone through TDH-39 earphones in MX 41/AR cushions. The same equipment was used for the hearing tests. Frequency of stuttering was recorded by means of a Linemaster Model 491-SC foot pedal switch which activated a General Controls Model CE 600 AS 602, 115 V AC counter. A Rustak four-channel event recorder was employed for this purpose in some sessions. An Olympic stopwatch was used to measure time intervals.

Procedure

Preliminary tests were run on 18 nonstuttering *Ss* with normal hearing to determine the specific frequency to be used as an aversive stimulus. Pure tone threshold tests indicated that 3 out of 4 male *Ss* had reduced acuity in both ears at 6,000 cps ranging from 5 to 15 db. At 4,000 cps only 1 male *S* had a 15-db. unilateral loss. Discomfort levels were established for each of three frequencies; 3,000, 4,000, and 6,000 cps. These ranged from 75 db. to 108 db., an average of about 100 db. Two *Ss* were dropped at this point because no discomfort level could be established within the maximum output of the audiometer. The three frequencies were arranged in their six possible pairings, and presented to each *S* in random order so he would judge each pair twice. At intensities above the discomfort threshold, he was asked to state whether the first or second tone in a given pair was more comfortable. Four *Ss'* responses were patterned so that either the first or second stimulus presented was always the one chosen. The remaining 12 *Ss* consistently selected the 4,000-cps tone as being the more aversive (the other two tones were chosen only six times in 140 trials).

The experiment was conducted in a two-room suite. The *Ss* could see neither *E* nor the equipment. The ambient noise level, measured before each session, varied from 35 to 43 db., sound-pressure level, averaged 40 db.

Potential *Ss* were screened to eliminate those with hearing losses or with thresholds of discomfort above 108 db. at 4,000 cps (maximum output of the audiometer at this frequency). A 3-min. tone decay test was also performed to guard against fading of the stimulus tone during the escape procedure. Two volunteers were rejected because of bilateral hearing losses.

Each *S* was seen on three different occasions, with a minimum of 3 days between experimental sessions. The written instructions indicated that he was to read the assigned passage aloud in a natural manner, and informed him that he would hear loud tones through the earphones during portions of the reading.

Headphones were worn throughout each session. The reading material consisted of three passages taken from Einstein's (1949) *The World As I See It*.

The base-line period varied in length from 6 to 25 min., 15 being about the average. When a stable base rate of stuttering frequency was established, after an initial period of adaptation, the conditioning period was initiated. During this 20-min. interval *S* was presented with one of the following: (a) tone contingent upon stuttering (aversive condition); (b) cessation of a continuous tone contingent upon stuttering (escape condition); (c) random presentation of tone (random condition).

In the extinction period *S* continued to read for another 20 min. during which time no further stimuli were presented. Throughout the base-line period, *E* recorded the rate of stuttering by pressing a foot pedal switch which operated an electric counter. The *S's* score was the number of stuttered words per passage read, following one of the procedures used by Flanagan et al. (1958). During the aversive conditioning period, the manual switch on the audiometer was pressed simultaneously, delivering a 1-sec. tone burst each time *S* stuttered. In the escape condition, an interrupter switch was used so the continuous tone stopped for 5 sec. each time a disfluency occurred. If *S* stuttered during this silent period, resumption of the tone was postponed for another 5 sec. In order to permit a comparison between the aversive and random conditions, schedules for all possible rates of stuttering were constructed randomly (Hodgman, 1945). The base rate of stuttering was calculated for a 5-min. interval at the end of the preconditioning period, and stimuli in the conditioning period were delivered according to the appropriate schedule. Since it was necessary for *E* to employ a stopwatch to time the scheduled intervals between stimulus presentations, frequency counts were made by one of two other observers, both members of the Psychology Speech Clinic Staff. The observers also performed this function during other sessions. They were not informed as to the nature of any given session, nor when their judgments would actually be used for scoring. When a four-channel event recorder was made available to *E*, it was possible to read correlations of judgments from the tapes—.88 in the case of one observer and .92 in the case of the other. This is sufficiently high that it may be assumed that scores derived from the

observers' frequency counts were essentially the same as those derived from those of *E*, that *E*'s judgments during the experiment were accurate, and the appropriate stimuli were delivered to *Ss*.

At the end of the third session each *S* was given a questionnaire in which he was asked to state his impressions of what the experiment might have been investigating, whether the stimulus was related to his stuttering, and what adjectives he would use to describe the stimulus (Table 4).

RESULTS

The essential finding of this study was that stuttering decreased in all conditions. The adaptation curves obtained on *Ss* (Figures 1 and 2 are typical) are all similar, regardless of the imposed experimental condition. This suggests that there is no treatment effect, which is exactly what an analysis of variance of the data proves (Table 1). This analysis was derived from a ratio of the main frequency of stuttering during the conditioning or extinction period divided by the average of the last two scores in the base-line period. A square-root transformation was applied to correct for nonhomogeneity of variance. Differences among experimental conditions fell short of significance ($p = .10$).

In the computation of the sums of squares, the sum of means for the aversive condition was 20.6, as compared with 17.6 for both the random and escape conditions. This indicates the direction of the trend: the random

and escape procedures resulted in slightly more response decrement than the punishment procedure.

To minimize the magnitude of *S* differences, a second analysis was carried out (Table 2). Scores were derived from the ratio ($x = b - f/b$, where *b* refers to the base-line average, and *f* to the mean number of stutterings in the conditioning or extinction period. This procedure reduced the *f* value for *Ss* from 247.311 to 7.220. However, it also reduced the significance of the effect of experimental conditions to the .25 level. The order in which the conditions were presented had no significant effect on performance.

TABLE 2
ANALYSIS OF VARIANCE OF RATIO DATA

Source	df	MS	f
Condition	2		
Conditioning		.0395	2.820*
Extinction		.0162	.474
<i>Ss</i>	5		
Conditioning		.1012	7.220***
Extinction		.0965	2.822**
Order	2		
Conditioning		.0020	.142
Extinction		.0057	.167
Error	8		
Conditioning		.0140	
Extinction		.0342	

Note.—The scores for this analysis were derived from a ratio: average base-line frequency minus the mean frequency of stuttering during the conditioning or extinction period/base line.

* $p < .25$.

** $p < .1$.

*** $p < .01$.

TABLE 1
ANALYSIS OF VARIANCE OF TRANSFORMED
RAW DATA

Source	df	MS	f
Condition	2		
Conditioning		.4797	4.096*
Extinction		.2009	.055
<i>Ss</i>	5		
Conditioning		28.9601	247.311**
Extinction		25.3515	69.209**
Order	2		
Conditioning		.1860	1.588
Extinction		.1452	.040
Error			
Conditioning		.1171	
Extinction		.3663	

Note.—The scores for this analysis were derived from a ratio of the mean frequency of stuttering during the conditioning or extinction period/average base-line frequency. A square-root transformation was applied to correct for nonhomogeneity of variance.

* $p < .1$.

** $p < .01$.

Table 3 shows the mean frequency of stuttering for all *Ss* in all conditions. When cumulative frequency graphs were plotted (as in Figures 1 and 2), suppression of response was notable in all conditions in all *Ss* with the exception of a rise in stuttering frequency in the aversive condition for *S* 3 (Figure 2). Curves for the random and aversive conditions were generally parallel, although response suppression was greater in the random condition for *Ss* 5 and 6. The rate of stuttering continued to drop during the extinction period except in three instances: *S* 2 following the random period, *S* 4 following the escape period, and *S* 5 following the aversive stimulus presentation. *S* 1 stuttered so severely that

it was rare for him to read for any 5-sec. interval without stuttering. Therefore, he experienced the escape condition as an occasional presentation of brief tones, and generalizations about the escape period are made only with reference to the other five Ss.

The cumulative frequency curves derived for the aversive condition appear similar to those presented by Flanagan et al. (1958). However, the curves for the escape periods in the two studies differ. In Figures 1 and 2

TABLE 3

MEAN FREQUENCY OF STUTTERING DURING BASE LINE, CONDITIONING, AND EXTINCTION PERIODS

Periods	Condition		
	Aversive ^a	Escape ^b	Random ^c
S 1			
Base-rate average	111.50	95.00	87.00
Conditioning mean	104.00	78.33	83.67
Extinction mean	90.00	60.33	72.67
S 2			
Base-rate average	8.00	5.00	10.50
Conditioning mean	4.38	2.33	4.67
Extinction mean	2.00	1.75	6.25
S 3			
Base-rate average	6.50	8.00	9.00
Conditioning mean	7.89	3.88	3.09
Extinction mean	4.00	3.22	2.67
S 4			
Base-rate average	2.50	3.00	3.50
Conditioning mean	1.10	.89	1.45
Extinction mean	1.00	.67	1.60
S 5			
Base-rate average	7.00	6.50	5.50
Conditioning mean	4.50	4.00	2.89
Extinction mean	4.70	2.00	1.22
S 6			
Base-rate average	9.00	10.50	8.00
Conditioning mean	5.57	5.71	3.00
Extinction mean	1.43	3.83	1.00

^a Tone contingent upon stuttering response.

^b Cessation of tone contingent upon stuttering response.

^c Tone independent of stuttering response.

it is apparent that the stuttering response is suppressed. This was true of all Ss in this condition (Table 3). The opposite result is evident in Flanagan, Goldiamond, and Azrin's (1958, Figure 1, p. 174) curves for S 1 and S 2. Their graph for S 3 is difficult to interpret, and it is not really clear whether the mean number of disfluencies for the conditioning period would be greater than a base-line average. Figure 1, in the present study, shows a curve for the aversive condition which is similar in shape to the Flanagan,

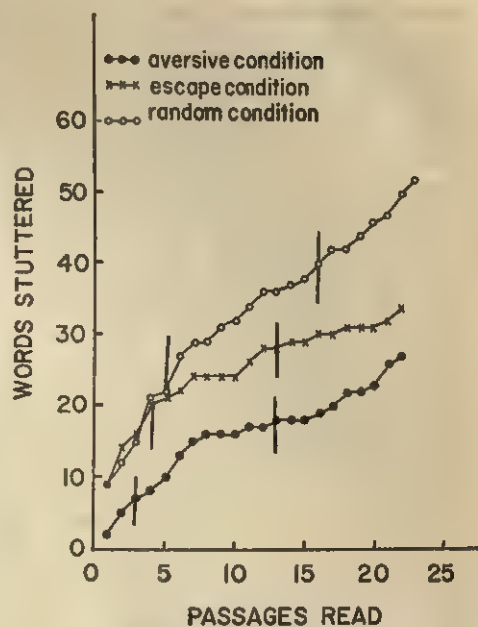


FIG. 1. Cumulative number of words stuttered during successive readings of the same passage—S 4. (Vertical lines indicate the boundaries of the conditioning period.)

Goldiamond, and Azrin graph for S 3, but the rise from base-line average only amounts to a difference of 1.4 disfluencies per passage.

Adaptation occurred as previously described in the literature, the greatest decrement usually occurring during the second reading, approaching an asymptote by the end of the third or fourth (Cullihan, 1963; Sheehan, 1951; Trotter, 1955). Two exceptions to this pattern occur in the base-line period for the random condition. Ss 1 and 6 both increased their stuttering initially to approach the base-line average of the other two conditions.

However, with two exceptions, the aversive condition for S 3 (Figure 2) and the random condition for S 4 (Figure 1), the introduction of the stimulus always resulted in an immediate drop in stuttering frequency, although the number of passages read during the base-line period varied from four to seven.

The Ss' ratings of the stimulus tone appear in Table 4. Although a 5-point scale was presented to Ss, no adjectives were rated with a 4 (very) or 5 (extremely). The differences between ratings are not large enough to indi-

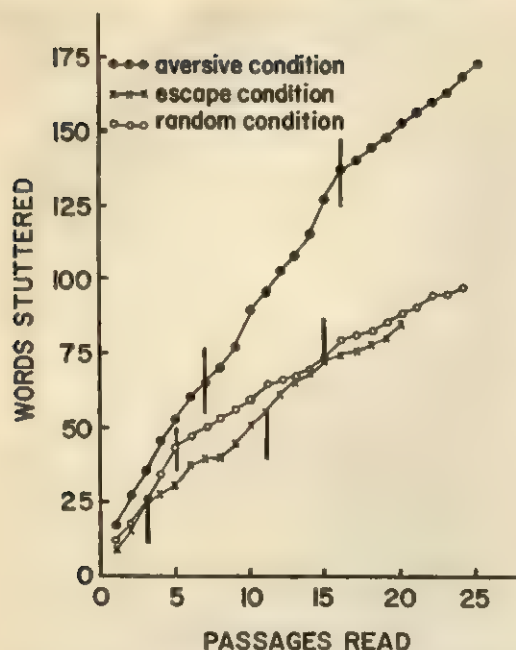


FIG. 2. Cumulative number of words stuttered during successive readings of the same passage—S 3. (Vertical lines indicate the boundaries of the conditioning period.)

cate more than a tendency to consider the stimulus as being more uncomfortable and distracting than painful or noxious. Only two of the six Ss indicated in the questionnaire that they had correctly interpreted the nature of the experiment and the relation of the stimulus to their disfluencies. Two Ss indicated that they began to ignore the tone after a time which raises the question of Ss adapting to the stimulus.

TABLE 4
SUBJECTS' RATINGS OF THE STIMULUS

Adjective	Subjects						
	1	2	3	4	5	6	M
Painful	0	1	1	2	0	0	.66
Noxious	0	2	2	2	0	0	1.00
Distracting	2	2	2	2	2	3	2.16
Irritating	2	3	2	2	1	2	2.00
Uncomfortable	3	3	2	2	0	3	2.16
Annoying	1	2	2	2	2	2	1.83
Punishing	0	2	1	2	0	2	1.16
Unpleasant	2	2	2	2	2	2	2.00

Note.—All Ss were asked to rate on the following scale: 1 = slightly, 2 = somewhat, 3 = quite, 4 = very, 5 = extremely.

DISCUSSION

The findings in the present study were similar to those of Flanagan, Goldiamond, and Azrin (1958) when a punishment procedure was employed, but differences were apparent in the results obtained from the use of an escape procedure. It is difficult to assess the magnitude of these differences since Flanagan et al. (1958) do not report numerical data. Assuming that the discrepancies are significant, they may be explained in terms of individual differences among Ss and procedural differences.

Since the effect of reading material was not controlled by Flanagan, Goldiamond, and Azrin, it cannot be analyzed. This leaves open to question the possibility that the pages read during periods of increased disfluency were more difficult phonetically, grammatically, or emotionally. The curves for two of the three Ss in this condition (Flanagan et al., 1958, p. 174) are not as smooth as the four other curves presented, an observation which might support this possibility. Order was not a significant factor in the present study, but the time period between trials was never less than 3 days. Flanagan, Goldiamond, and Azrin (1958) ran their Ss on consecutive days, so there may have been some carry-over from one trial to the next. Their S 2 and S 3 (Flanagan et al., 1958, p. 174), who experienced the escape condition in the second trial, show less obvious response increment than S 1, who encountered this condition in his initial trial.

The statistical analyses of data from the present study indicated that the three operant conditioning techniques employed had essentially the same effect: all resulted in suppression of the stuttering response with no significant difference in the amount of suppression. The only trend which could be noted indicates that the contingent presentation of the aversive stimulus resulted in less suppression of the stuttering response than noncontingent, or random, presentation. Since conservative statistical procedures were employed, this trend might manifest itself more strongly if a larger number of Ss were studied.

Before the study was undertaken, it appeared that the 4,000-cycle tone at 108 db.

would be effective as a noxious stimulus, since there was reason to believe that it was more aversive than the tone employed by Flanagan, Goldiamond, and Azrin. First, Ss in the pretest had judged it to be so. Second, although all of the stutterers in the experiment had "normal" hearing, five of the six had a drop in acuity of 5 to 30 db. between 4,000 and 6,000 cps, a factor which would certainly have reduced the experienced discomfort had the 6,000-cycle tone been employed. Flanagan, Goldiamond, and Azrin did not test the hearing of their Ss before their experiment. Third, there was a difference of 20 db. in the noise levels present in the experimental rooms. Hirsch (1952) reports that when the effect of earphone cushions is taken into account (cutting out about 20 db. of the external noise), an ambient noise level of 40 or 50 db. may be allowed "without anticipating measurable effects on the absolute thresholds [p. 166]." The noise level reported here never exceeded 43 db., so it may be assumed that the stimulus was received at the full 108-db. intensity level. Flanagan, Goldiamond, and Azrin (1958) report the presence of a constant 60-db. noise. This could have masked their stimulus tone so that its *effective* level was 10 db. less than the 105 db. reported.

Despite this, the stimulus employed always had the same effect on behavior, whether its application or removal was contingent on stuttering or whether its presentation was random. In all conditions stuttering frequency was reduced. Thus it is apparent that there was, in fact, only one experimental condition: a 4,000-cps tone at 108 db. whose introduction into the environment was followed, almost invariably, by a drop in the established rate of stuttering. Since cessation of the tone contingent upon stuttering did not increase the frequency of nonfluent behavior, the stimulus was not functionally aversive although it was rated as subjectively unpleasant.

If the word "punishment" is used in a broader, non-Skinnerian context, it is possible to interpret the results as an indication that a mild punishment may be effective in reducing symptomatic behavior. In future experimentation it might be desirable to in-

clude a control group receiving a neutral tone for the supposedly noxious tone, to separate more clearly the roles of distraction and punishment. However, previous studies by Barber (1939), Shane (1955), and Cherry and Sayers (1956) provide ample evidence that the introduction of a novel auditory stimulus does in itself provide sufficient distraction to reduce stuttering behavior even though the stimulus is not aversive. The present findings are clearly in line with these previous studies but not in line with those of Flanagan, Goldiamond, and Azrin.

Since the stimulus employed in this expanded replication was independently rated as only mildly aversive, and did not produce contingent effects, the parsimonious explanation appears to be that the tone operated in general as a distraction device.

REFERENCES

- BARBER, V. Studies in the psychology of stuttering: Chorus reading as distraction. *Journal of Speech Disorders*, 1939, 4, 371-383.
- BLOODSTEIN, O. Hypothetical conditions under which stuttering is reduced or absent. *Journal of Speech and Hearing Disorders*, 1950, 15, 142-143.
- CHEERY, C., & SAYERS, B. Experiments upon the total inhibition of stammering by external control, and some clinical results. *Journal of Psychology*, 1956, 1, 233-246.
- CULLINAN, W. L. Stability of consistency measures in stuttering. *Journal of Speech and Hearing Research*, 1963, 6, 134-138.
- EINSTEIN, A. *The world as I see it*. New York: Philosophical Library, 1949.
- FLANAGAN, B., GOLDDIAMOND, I., & AZRIN, N. Operant stuttering: The control of stuttering behavior through response contingent consequences. *Journal of the Experimental Analysis of Behavior*, 1958, 1, 173-177.
- FREDERICK, C. J. An investigation of learning theory and reinforcement as related to stuttering behavior. Unpublished doctoral dissertation, University of California, Los Angeles, 1955.
- FRICK, J. V. An exploratory study of the effect of punishment (electric shock) upon stuttering behavior. Unpublished doctoral dissertation, State University of Iowa, 1951.
- HILL, H. E. An experimental study of disorganization of speech and manual response in normal subjects. *Journal of Speech and Hearing Disorders*, 1954, 19, 295-305.
- HIRSCH, I. J. *The measurement of hearing*. New York: McGraw-Hill, 1952.
- HODGMAN, C. D. (Ed) *C.R.C. standard mathematical tables*. Cleveland: Chemical Rubber, 1945.

- JOHNSON, W. Introduction: The six men and the stuttering. In J. Eisenson (Ed.), *Stuttering: A symposium*. New York: Harper & Row, 1958.
- SHAMES, G. H., & SHERRICK, C. E., JR. A discussion of nonfluency and stuttering as operant behavior. *Journal of Speech and Hearing Disorders*, 1963, 28, 3-18.
- SHANE, M. L. S. Effect on stuttering of alteration in auditory feedback. In W. Johnson (Ed.), *Stuttering in children and adults*. Minneapolis: University of Minnesota Press, 1955.
- SHEEHAN, J. G. The modification of stuttering through non-reinforcement. *Journal of Abnormal and Social Psychology*, 1951, 46, 51-63.
- SHEEHAN, J. G. Theory and treatment of stuttering as an approach-avoidance conflict. *Journal of Psychology*, 1953, 36, 27-49.
- SHEEHAN, J. G. Conflict theory of stuttering. In J. Eisenson (Ed.), *Stuttering: A symposium*. New York: Harper & Row, 1958.
- SHEEHAN, J. G., CORTESE, P., & HADLEY, R. Guilt, shame and tension in graphic projections of stuttering. *Journal of Speech and Hearing Disorders*, 1962, 27, 129-139.
- SHEEHAN, J. G., & VOAS, R. B. Stuttering as conflict: I. Comparison of therapy techniques involving approach and avoidance. *Journal of Speech and Hearing Disorders*, 1957, 22, 714-723.
- SIEGEL, G. M., & MARTIN, R. R. Experimental modification of disfluency in normal speakers. *Journal of Speech and Hearing Research*, 1965, 8, 235-245. (■)
- SIEGEL, G. M., & MARTIN, R. R. Verbal punishment of disfluencies in normal speakers. *Journal of Speech and Hearing Research*, 1965, 8, 245-251. (b)
- SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- STASSI, E. Disfluency of normal speakers and reinforcement. *Journal of Speech and Hearing Research*, 1961, 4, 358-361.
- TAYLOR, I. K. What words are stuttered? *Psychological Bulletin*, 1966, 65, 233-242.
- TROTTER, W. D. The severity of stuttering during successive readings of the same material. *Journal of Speech and Hearing Disorders*, 1955, 20, 17-25.
- VAN RIVER, C. The effect of penalty upon frequency of stuttering spasms. *Journal of Genetic Psychology*, 1937, 50, 193-195.

(Received November 13, 1967)

SELF-DIRECTED PROGRAM FOR WEIGHT CONTROL: A PILOT STUDY

MARY B. HARRIS¹

Stanford University

A treatment program was designed to enable Ss to lose weight through the use of self-monitored techniques for changing their eating behaviors. All Ss who participated in the program achieved a stable loss in weight, and their mean loss was significantly greater than the change shown by a group of similarly motivated control Ss. No additional effects due to a few sessions of aversive counterconditioning were demonstrated, and no general mood changes accompanied the weight loss. The Ss did report a decreased temptation to overeat. It was suggested that similar programs of gradual habit change through self-control of stimulus conditions and reinforcement contingencies might be applied to the treatment of other addictive behaviors, which are also very refractory to change.

Attempts to deal with the undesirable behavior of overeating have been spectacularly unsuccessful. Although a great number of causal agents or correlates have been postulated, such as depression (Simon, 1963), anxiety (Cauffman & Pauley, 1961), power orientation (Suczek, 1957), a variety of other personality problems (Bruch, 1957; Kaplan & Kaplan, 1957, who list 28 suggested meanings of obesity; Shipman & Plesset, 1963), insufficient exercise (Mayer, 1955), presence of night-eating syndrome (Stunkard, 1959a), lack of correlation between report of hunger and gastric motility (Stunkard, 1959b), obesity of parents (Cappon, 1958), more dependence on external stimuli such as flavors and time of day to regulate hunger (Schachter, 1967), and no doubt many others, no clear causal relationships have been substantially detailed. An equally large variety of treatments have been tried, ranging from such nonacademic means as low-calorie products, exercise salons, yoga, and reducing clubs to hypnosis (Erickson, 1960), dietary instruction (Young, Moore, Berresford, Einset, & Waldner, 1955), appetite depressants and other drugs (Silverstone & Solomon, 1965), general medical advice (Stunkard,

1958), psychoanalysis and other forms of psychotherapy (Bruch, 1957), group discussion of physical and emotional factors (Harmon, Purkonen, & Rasmussen, 1958), aversion-relief therapy (Thorpe, Schmidt, Brown, & Castell, 1964), recording all eating (Ferster, Nurnberger, & Levitt, 1962; Stollak, 1966), self-control (Ferster et al., 1962; Goldiamond, 1965), aversive counterconditioning with shock (Stollak, 1966; Wolpe, 1958), operant conditioning with shock (Meyer & Crisp, 1964), and aversive counterconditioning with nausea (Cautela, 1966).

In general, the reported attempts to treat overweight people fall into two categories: case histories of techniques which proved successful with one or a very few patients (Cautela, 1966; Erickson, 1960; Ferster et al., 1962; Goldiamond, 1965; Meyer & Crisp, 1964; Thorpe et al., 1964; Wolpe, 1958) and survey studies of patients in a medical setting, most of which have reported a general lack of success in effecting any long-range weight reduction (Franklin & Rynearson, 1960; Harmon et al., 1958; Shipman & Plesset, 1963; Silverstone & Solomon, 1965; Young et al., 1955). An experimental study by Stollak, using aversive counterconditioning therapy with electric shock and adequate control groups, reported the same poor results (a mean net loss of less than 5 lb.). The inadequacy of treatment has been nicely summarized by Stunkard (1958)—“most obese persons will not stay in treatment for obesity.

¹ The author wishes to express her appreciation to Albert Bandura for his generous assistance with all phases of the study. Request for reprints should be to the author, Department of Educational Foundations, University of New Mexico, Albuquerque, New Mexico 87106.

Of those who stay in treatment, most will not lose weight and of those who do lose weight, most will regain it [p. 79]." The evidence to disconfirm this description has not yet appeared.

An examination of the contingencies governing addictive behavior in general and overeating in particular shows several reasons why this type of behavior should be so resistant to change. Addictive behaviors such as overeating provide immediate positive reinforcement for the individual, while the reinforcement for refraining from eating is usually extremely delayed. Moreover, the aversive consequences of overeating are typically delayed for weeks or even years. As Eysenck (1960) has pointed out, therapies based on pairing aversive reinforcement with the performance of the undesirable behavior are likely to be ineffective in the long run, because the fear of the negative consequences will tend to be extinguished when the behavior is performed without the negative reinforcement following. In addition, there are two other variables which make overeating behavior particularly resistant to alteration. Both Ferster et al. (1962) and Goldiamond (1965) have pointed out that eating behavior occurs in a very wide range of situations and is under the control of many stimuli other than those physiological ones causing hunger. A second factor, one unique to overeating among the addictions, is that one cannot be delivered from temptation. Drugs, cigarettes, sexual perversions, and alcohol can all be given up completely and the stimuli associated with them can be avoided; however, everyone must eat at least two or three times a day, and it is impossible to avoid exposure to situations and performance of behaviors one wishes to avoid. Thus the control of eating shares elements with the control of other undesirable appetitive behaviors but is further complicated by the facts that the stimuli for eating are ubiquitous and that some eating behaviors must be performed several times a day.

The study to be discussed attempted to use some of the specific procedures and methods of analysis which have been found useful in controlling both overeating and other undesirable behaviors. Many of these techniques,

particularly those of Ferster et al. (1962) and Goldiamond (1965) place a very strong emphasis on the development of self-control through altering the stimulus conditions under which the behavior occurs and generating self-produced consequences for the behavior. A second major influence was the use of aversive conditioning, both operant (Meyer & Crisp, 1964) and classical (Stollak, 1966; Wolpe, 1958), particularly the nausea conditioning of Cautela (1966). For the present study, nausea has several advantages over electric shock. It is completely under the individual's control, can be used in any situation, and is a response directly antagonistic to eating. In addition, a group setting was used in this research, primarily because of convenience, but also to take advantage of any slight increase in motivation which group discussion and support might generate.

Because long-term control of eating has been so refractory to change, it was felt that there was little to be gained by employing an attention-placebo control group in addition to a no-treatment control group. For the same reason it was decided to study the efficacy of a treatment approach combining several different components rather than comparing different techniques. The study was thus designed as a broad program to produce change by use of all the procedures which might assist an individual in attaining control of his own eating behavior by gradually approximating the eating pattern he eventually wanted to maintain. Drugs, lists of forbidden foods, or rigid rules of only three meals per day were not used, as they were felt to be useless in the maintenance of permanent habits. Although the procedures used were often geared to the individual's specific problems with eating, the type of approach should prove amenable to extension to dealing with other types of addictive behaviors.

METHOD

Subjects. The Ss for this study were both men and women who had answered an advertisement in the Stanford University newspaper, who were at least 15 lb. overweight, and who were not being medically treated for overweight or any serious illness. Two individuals who spontaneously began to discuss their psychiatric problems in a 10-min. interview with E were not included in the study. From the pool of

satisfactory Ss, two sections of three men and five women each who had the same available time for meetings were selected, along with a random group of three men and five women from the remaining potential Ss to serve as a control group.

Control group. The members of the control group were truthfully told that the experimental groups would not be meeting during their available hours. They were then asked if they would attempt to lose weight on their own and if they would consent to be reweighed at the conclusion of the experiment. All Ss agreed. They were then weighed without shoes on a commercial bathroom scale, to the nearest half pound, given a calorie chart, and reminded that the only way to lose weight was to change their eating habits permanently.

Treatment procedure. The two experimental sections of eight Ss each met with E twice a week for approximately 2 mo. until the end of the school quarter. At the beginning of every meeting Ss were weighed individually without shoes on a commercial bathroom scale to the nearest half pound. The Ss were all given calorie charts and asked to keep a record of their normal eating habits for 1 wk. and of their daily intake, including place and time, for the next several months. At the first meeting and approximately every 3 wk. thereafter Ss completed a brief questionnaire consisting of eight rating scales. Four of the scales concerned their levels of tension and depression both generally and with specific reference to eating; three scales concerned temptations to eat rich foods, large meals, and snacks; and one concerned time thinking about food. One purpose of this questionnaire was to ascertain whether any mood changes would accompany or result from the process of weight reduction, as is suggested by such theories as "obesity as a depressive equivalent" (Simon, 1963). Another purpose of the rating scale was to see whether changes in the temptation to eat would accompany changes in eating behavior.

Three very general aspects of the program were discussed thoroughly with Ss. It was stressed that the program was designed to enable each individual to develop permanent eating habits he could maintain indefinitely. For this reason no arbitrary restrictions were placed on diet, and the individual was urged to consider what type of permanent eating pattern he would like to maintain. Along with the emphasis on establishing permanent habits there was an emphasis on awareness of what one was eating, especially caloric content and food value, and of the reasons for eating and the situation in which the eating occurred. A third general aspect of the treatment program was its emphasis on making only gradual changes. It was felt that making drastic changes, although perhaps producing greater immediate loss, would make it too difficult to maintain the changes indefinitely. The Ss were urged to make only small changes which they were sure they could maintain and which were in the direction of the permanent eating pattern they desired to establish.

Techniques. Three general types of techniques for approaching their goal of altered eating habits were

recommended to Ss during the first 2 mo. of the program. First, Ss were told about the concepts of positive and negative reinforcement, and it was explained to them how eating behavior was controlled by immediate positive reinforcement and only very delayed negative reinforcement. All Ss were asked to consider the positive outcomes maintaining their eating and to hand in a list of the aversive consequences causing their desire to lose weight. In group discussions suggestions were made as to methods useful for making reward for refraining from eating more immediate and for making aversive consequences come immediately after beginning to eat; among the suggestions were the reciting of the list of reasons for losing weight when tempted to overeat, the viewing of an unattractive picture of oneself in a bathing suit when tempted, self-rewards of money, movies, clothes, etc. for every day or week of good eating habits or for every pound lost. It was emphasized that food should not be used as a reward.

The second group of techniques concerned the stimulus control of eating behavior and attempted to limit the situations which were associated with eating. The Ss were asked to analyze the stimuli governing their eating behavior and were given suggestions about limiting the number of situations in which they eat and of stimuli connected with eating. Among these suggestions were many concerning limitations of place and time of eating—for example, try to eat sitting in only one chair at the same table or in as few places as possible; never eat while standing or in the office, bedroom, living room, etc.; try to eat meals and planned snacks only at certain definite times during the day. Other suggestions concerned the limitation of the availability of other reinforcements for eating and of tempting foods—for example, don't eat while reading, watching television, studying, etc.; try to avoid walking past candy machines, snack bars, etc. The Ss were also informed of some Schachter's research (Schachter, 1967) demonstrating the control of eating behavior by external rather than internal stimuli in overweight people.

The third general group of suggestions dealt with the actual behaviors involved in eating. It was pointed out that many overweight people, particularly people on diets, eat much more rapidly than people of normal weight, and that it takes at least 15 min. after beginning to eat before one begins to feel the effects of the food. It was thus suggested that Ss try to break up the chain of eating by not putting any more food on the fork until they had finished chewing and swallowing the last bite, that they chew their food very slowly and with great attention to taste and texture like a gourmet, that they practice taking short breaks of not eating during a meal, and that they begin to try out all of these suggestions only near the end of a meal when they are not so hungry that these behaviors would become aversive. It was also recommended that they attempt to practice leaving a little bit of every

food on their plate, rather than automatically eating it all.

Throughout the presentation and discussion of all these techniques for controlling eating habits, it was emphasized that these were all methods for self-control of these behaviors. Although each technique should prove to be useful in some cases, it was recognized that certain techniques would be differentially useful for certain individuals; for this reason, each *S* was asked to try out all techniques which he felt might be appropriate for regulating his own behaviors and to try to commit himself to one small change at a time in the direction of his permanent eating habits. After the above meetings were completed, so that *Ss* had learned principles and techniques of reinforcement and self-control of eating behaviors, they were given a paper summarizing these principles and techniques. A list of low calorie foods and beverages suggested for cooking and eating was also given to them.

Two additional aspects of the treatment procedure were the discussion of nutrition and training in relaxation. A nutritionist spoke to *Ss* about diet in general, answered many questions about the nutritional and caloric values of various foods, and made brief verbal and written comments to *Ss* about the nutritional adequacy of their individual eating records. The *Ss* were also given training in relaxation exercises by listening to a tape and reading a manual designed by *E* to accompany the tape. The relaxation exercises were presented as an alternative to eating in dealing with tension, depression, and boredom.

Differential treatment of subgroups. After a period of approximately 2½ mo. and following a change of schedule for those *Ss* who were university students, each of the two earlier experimental sections was subdivided into an aversive conditioning subgroup and a continuation subgroup. As two participants had withdrawn from the program, only the remaining 14 *Ss* were subdivided. The division was made into two subgroups such that each had 4 from one original section and 3 from the other, each had three males and four females, three good friends were put in one subgroup, each had about the same number of those who showed regular attendance in earlier meetings, and each had approximately the same mean weight loss for men and for women up to that time. A random selection was then made as to which of the two subgroups would be designated the continuation subgroup and which the aversive counterconditioning subgroup. Two hours per week were chosen at times when the continuation *Ss* were free, and these *Ss* were told that they could attend group meetings at one or both of these times. At these meetings, discussion centered around current problems *Ss* were having with eating and included suggestions and support from the other group members. The aversive counterconditioning *Ss*, on the other hand, did not have further group meetings, although a few of these *Ss* attended one or two of the continuation subgroup meetings. Instead, the aversive counterconditioning was carried out in indi-

vidual meetings with *E*, which were held approximately weekly.

Aversive counterconditioning. The procedures used in conditioning were based on a report by Cautela (1966). The *Ss* who were not able to relax sufficiently after the group instructions in relaxation were given further individual instruction until they were able to experience a moderate degree of relaxation. Each *E* was then given a thorough explanation of and rationale for aversive counterconditioning including the fact that the response was completely under his control and that the nausea conditioning would only be applied to certain foods in certain situations in which he was sure he did not want to eat. As work with two *Ss* in a preliminary study had demonstrated that nausea conditioning could also be effectively applied to reduce the positive valence of such behaviors as eating all the food on ones plate or eating more than one cookie, these were also mentioned as possibilities. Each *S* then chose a specific food in a specific situation where he desired to refrain from eating and was asked to describe the situation to *E* in great detail. The *S* was also asked to recall the last time he felt nauseous and vomited and to describe to *E* the sensations associated with nausea. Surprisingly, these varied a great deal among *Ss*.

During the actual nausea conditioning session, *S* sat comfortably in a chair with eyes closed and tried to visualize and imagine as vividly as possible the scene *E* described to him. The *E* then described in detail the situation *S* had chosen and described *S's* eating the food, feeling more and more nauseous and sick with every bite, and then vomiting. On alternate trials, *S* was described as about to eat the food, realizing how he didn't really want it, deciding not to eat the food, and then relaxing and feeling very happy. All but one *S* reported that their imagery and feelings of nausea were extremely vivid. One *S* (who had lost 20 lb.) reported at the next meeting that he did not want to undergo nausea conditioning again, because he had found it unpleasant. All the other *Ss* reported that they felt the experience was very effective. Usually two aversive and two relief trials were given in each session. Some *Ss* chose different situations for each aversive counterconditioning session; some practiced aversive counterconditioning at home; some did neither or both. Due to the approaching end of the academic year and because some *Ss* required lengthier instruction in relaxation exercises than others and some were unable to meet with *E* each week, each *S* had only one to three sessions of aversive counterconditioning.

Final weighing. A final meeting was held approximately 4 mo. after the study began in order to get a final weight for all experimental and control *Ss*. All *Ss* who could not attend the meeting were contacted and final weights were obtained for all but the two *Ss* who had previously stopped participating and one control *S* who repeatedly broke appointments and finally could not be located. At the time of the final weighing, the experimental program was

discussed with control Ss, and they were given copies of the paper on suggestions for weight reduction, the list of low calorie foods, and the manual of relaxation exercises.

RESULTS

The weights for individual experimental and control Ss are reported in Table 1. All 14 Ss who remained in the study as well as the 2 who dropped out lost weight, with the difference between pretest and posttest weights being highly significant ($t = 5.19$, $p < .001$). These losses are represented graphically in Figure 1.

Tables 2 and 3 show the results of analyses of variance for pounds lost and percentage weight loss, respectively, for all experimental Ss versus the controls. The effect of the experimental treatment is highly significant us-

ing both measures ($p < .001$), and the effect of sex is significant at the $p < .05$ level for pounds lost and at the $.05 < p < .10$ level for percentage weight loss. Clearly, the sample is too small to allow definite conclusions to be reached about whether men are likely to lose more weight than women independent of original weight. A t test on the difference between the additional weight losses of the aversive conditioning and continuation Ss during the last month and a half, when aversive conditioning Ss were receiving the extra treatment, revealed no significant differences ($t = .201$) between the two experimental subgroups.

The questionnaire data generally did not show large changes in the mood indexes over time. Only the data from those 12 Ss who had taken the questionnaire both in late January

TABLE 1
WEIGHT LOSSES FOR INDIVIDUAL EXPERIMENTAL AND CONTROL SUBJECTS

	Original weight	Loss at 2½ mo.	Final weight	Total change	Percentage of change
Aversive counterconditioning subgroup					
Male	196.5	-15.5	176.0	-20.5	-10
Male	197.0	-20.5	171.0	-26.0	-13
Male	198.0	-3.5	194.5	-3.5	-2
Female	135.0	-3.5	124.5	-10.5	-8
Female	143.0	-2.5	137.5	-5.5	-4
Female	159.0	-6.0	152.0	-7.0	-4
Female	135.0	-5.0	130.0	-5.0	-4
Mean for males	197.2	-13.2	180.5	-16.7	-8
Mean for females	143.0	-4.3	136.0	-7.0	-5
Mean for subgroup	166.2	-8.1	155.1	-11.1	-7
Continuation subgroup					
Male	190.0	-8.0	175.0	-14.5	-8
Male	218.5	-11.5	208.0	-10.5	-5
Male	214.0	-14.0	193.0	-21.0	-10
Female	155.5	-7.5	146.5	-9.0	-6
Female	143.0	+5	137.0	-6.0	-5
Female	152.0	-7.0	145.5	-6.5	-4
Female	129.5	-2.5	127.5	-2.0	-2
Mean for males	207.5	-11.2	192.2	-15.3	-7
Mean for females	145.0	-4.1	139.1	-5.9	-4
Mean for subgroup	171.7	-6.9	161.9	-9.9	-6
Mean for treatment males	202.4	-12.2	186.4	-16.0	-8
Mean for treatment females	144.0	-4.3	137.6	-6.5	-5
Mean for treatment group	169.0	-7.5	158.5	-10.5	-6
Control Ss					
Male	181.0		179.5	-1.5	-1
Male	178.5		166.5	-12.0	-7
Male	238.0		254.5	16.5	7
Female	149.0		144.0	-5.0	-3
Female	147.0		159.0	12.0	8
Female	143.0		155.0	12.0	8
Female	146.5		150.0	3.5	2
Mean for control males	199.2		200.2	1.0	.5
Mean for control females	146.4		152.0	5.6	4
Mean for control group	169.0		172.6	3.6	2

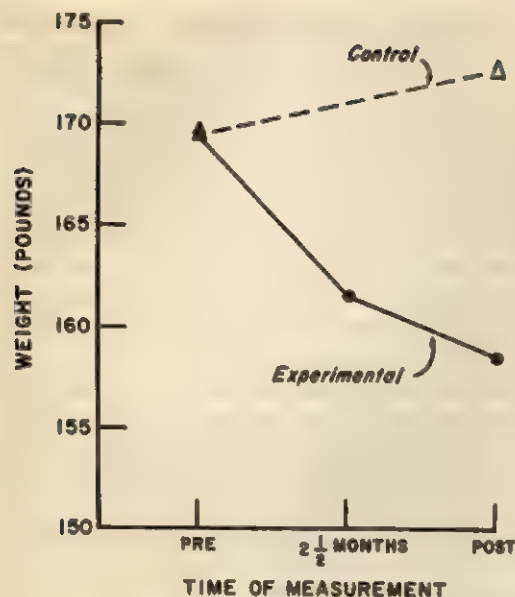


FIG. 1. Mean weights for experimental and control Ss.

and in mid April were analyzed, as fewer Ss completed questionnaires in May and June. General level of tension or depression as well as tension or depression specifically related to eating showed only nonsignificant decreases from beginning to end of the experiment and no relationship to amount of weight loss. Both the temptation to eat large meals ($t = 2.34$, $p < .05$) and a general index of temptation to overeat ($t = 2.22$, $p < .05$) showed significant decreases over the time course of the experiment, but the correlation between percentage weight loss and decrease in temptation was not significant ($r = .186$). The study thus provides no evidence that dieting necessarily causes large mood changes, although there is some evidence that this program can

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE
FOR POUNDS LOST

Source	SS	df	MS	F
Sex	322.32	1	322.32	5.51*
Treatment	938.15	1	938.15	16.04**
Interaction	27.86	1	27.86	<1
Within	994.41	17	58.49	
Total	2282.74	20		

* $p < .05$.
** $p < .001$.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE
FOR PERCENTAGE WEIGHT LOSS

Source	SS	df	MS	F
Sex	67.06	1	67.06	4.03*
Treatment	304.02	1	304.02	18.25**
Interaction	.58	1	.58	<1
Within	283.29	17	16.66	
Total	654.95	20		

* $.10 > p > .05$.

** $p < .001$.

reduce the temptation to overeat over a 3-mo. period.

DISCUSSION

The minimal requirement for a weight reduction routine is that it enable people to lose weight and keep it off for a moderate amount of time without causing serious disruption of the individual's daily life or health. More desirable attributes would be the production of permanent weight loss with minimal expense of time or money and with the new eating pattern becoming both habitual and satisfying, so that the individual is not constantly troubled by desires to eat. It is clear that the program described in this study does satisfy the minimal requirements. All Ss showed a decrease in weight, a loss which was significantly greater than zero and than the slight gain shown by control Ss. Expense in money was nil and time spent attending meetings varied from about 10 to 20 hr.

Several of the less essential but desirable features are also found in this program. All Ss expressed satisfaction with the program and stated that they had found it extremely helpful. The low dropout rate of 12.5% compares very favorably with those reported in the literature, such as 39% not returning after one visit (Stunkard, 1959c), 33% not answering questionnaires after 6 mo. (Franklin & Rynearson, 1960), 66% dropping out within 1 yr. (Silverstone & Solomon, 1965), 27% after one meeting (Shipman & Plesset, 1963). Of the Ss who received aversive conditioning, only one was not enthusiastic about the technique. This was the only instance of any S objecting to any aspect of the procedure; part of the success of the program

may have been that *Ss* were free to choose the techniques and eating patterns which suited them best rather than being forced to conform to rigid rules.

The questionnaire data show no significant mood changes as a result of the program, although the changes are in the direction of greater happiness and relaxation. The only changes which were significant were those indicating a lesser desire to overeat at meals and a decrease in general temptation to overeat at meals, eat sweets, and eat between meals. It seems likely that 3 mo. is not a sufficient time for large changes in affective values of eating to occur. It is also possible that questions dealing with eating behaviors rather than attitudes would reveal more significant changes.

The data appear to indicate that the program may be more effective for males than females, a difference also reported in a study by Stunkard (1959c). Although the possibility of physiological differences is not ruled out, the variables of awareness and motivation probably contribute to the sex differences found. In general, the men participating in this study were less aware of the caloric and nutritional values of various foods and of what their own eating habits were than were the women; moreover it is possible that the commitment to a group weight-reduction program represents a greater degree of motivation for men than for women, who often have a long history of going on diets with "the girls." A larger study will be necessary to discover the extent and causes of these sex differences, which were not quite significant when percentage weight lost rather than pounds lost was used as the measure.

The most crucial test of a weight reduction program or indeed of any therapy program is the permanence of its changes. Although long-term data are very difficult to obtain on members of a fluctuating university population, some data are available on *Ss* who participated in a preliminary study of many of the techniques used in this final program. Twelve female *Ss* originally began the pilot study; two dropped out very early, one for medical reasons and pregnancy. Eight of the *Ss* met with *E* in two groups; two of the *Ss*

met individually with *E* and received aversive conditioning. All *Ss* learned about stimulus control procedures. Each *S* kept daily eating records before and during the period of the experiment. Most of the *Ss* met regularly with *E* for only 2 to 3 mo., although follow-up data were gathered on all available *Ss* about 6-7 mo. after the start of the study. All but one *S* lost weight after 2-3 mo. with the mean short-term and also long-term loss for all pilot *Ss* being 6 lb. Thus it does appear, for this preliminary study, that some of the effects of this program do last for at least 6 or 7 mo. Other long-term data are provided by *E*, who lost about 27 lb., regained about 12 lb. and has lost about 7 of those, 12 mo. after the beginning of the original study. However, in view of the extremely poor long-term results of other weight reduction programs, it is clear that no conclusion can be drawn about the permanence of these changes.

It is possible that several variables may have contributed to the success of this particular study besides the planned experimental procedures. The modeling effect of *E*, who went from fat to moderate with the pretest *Ss* and from moderate to thin with *Ss* in the final study, was commented upon by many of the *Ss*. Many of the *Ss* stated that participating in a research project where data they provided were important increased their motivation. The majority of these *Ss*, moreover, were college educated, which may have made it easier for them to assimilate such concepts as reinforcement, stimulus control, and caloric values.

It is also possible that certain of the variables in this program, such as the group interaction, might make no additional contribution beyond that of the other procedures. It is possible that certain procedures, such as eating only when sitting down at the table, might be effective for all *Ss* whereas others, such as rewarding oneself monetarily for every day of good eating habits, might be of no use to any *S*. A much more controlled study in which various techniques and combinations of procedures are isolated would be necessary to discover their differential effects. In this study no additional effect of a very few sessions of aversive conditioning was found, although

most Ss reported that it had reduced or eliminated their desires for those foods and the two pretest Ss given aversive conditioning continued to lose weight at least over a period of 6 mo. It is conceivable that aversive conditioning, both classical and operant, will turn out to be a useful tool for Ss who do have cravings for certain foods or for eating everything set before them, although it may well be useless for Ss whose problem is lack of awareness of their eating habits and caloric values. One might expect the effects of aversive counterconditioning used as an adjunct to such a program as the one described in this study to be much more enduring than the effects of aversive counterconditioning sessions alone. In short, the study was designed to see if any program for long-term weight reduction through change of eating habits could be achieved; future studies will be necessary to discover the contributions of specific variables, and the permanence of the weight losses achieved.

REFERENCES

- BRUCH, H. *The importance of overweight*. New York: Norton, 1957.
- CAPPON, D. Obesity. *Canadian Medical Association Journal*, 1958, 79, 568-573.
- CAUFFMAN, W. J., & PAULEY, W. G. Obesity and emotional status. *Pennsylvania Medical Journal*, 1961, 64, 505-507.
- CAUTELA, J. R. Treatment of compulsive behavior by covert sensitization. *Psychological Record*, 1966, 16, 33-41.
- ERICKSON, M. A. The utilization of patient behavior in the hypnotherapy of obesity: Three case reports. *American Journal of Clinical Hypnosis*, 1960, 3, 112-116.
- EYSENCK, H. J. *Handbook of abnormal psychology*. New York: Basic Books, 1960.
- FERSTER, C. B., NURNBERGER, J. I., & LEVITT, E. B. The control of eating. *Journal of Mathetics*, 1962, 1, 87-109.
- FRANKLIN, R. E., & RYNEARSON, E. H. An evaluation of the effectiveness of dietary instruction for the obese. *Staff Meetings of the Mayo Clinic*, 1960, 35, 123-131.
- GOLDIAMOND, I. Self-control procedures in personal behavior problems. *Psychological Reports*, 1965, 17, 851-868.
- HARMON, A. R., PURKONEN, R. A., & RASMUSSEN, L. P. Obesity: A physical and emotional problem. *Nursing Outlook*, 1958, 6, 452-456.
- KAPLAN, H. I., & KAPLAN, H. S. The psychosomatic concept of obesity. *Journal of Nervous and Mental Disorders*, 1957, 125, 181-201.
- MAYER, J. Exercise does keep the weight down. *Atlantic Monthly*, 1955, 196, 63-66.
- MEYER, V., & CRISP, A. H. Aversion therapy in two cases of obesity. *Behavior Research and Therapy*, 1964, 2, 143-147.
- SCHACHTER, S. Obesity as a cognitive error. Speech presented at Stanford Medical School, February 14, 1967.
- SHIPMAN, W. G., & PLESSET, M. R. Anxiety and depression in obese dieters. *Archives of General Psychiatry*, 1963, 8, 530-535.
- SILVERSTONE, J. T., & SOLOMON, T. The long-term management of obesity in general practice. *British Journal of Clinical Practice*, 1965, 19, 395-398.
- SIMON, R. I. Obesity as a depressive equivalent. *American Medical Association Journal*, 1963, 183, 208-210.
- STOLLAK, G. E. Weight loss obtained under various experimental procedures. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1966.
- STUNKARD, A. J. The management of obesity. *New York State Journal of Medicine*, 1958, 58, 79-87.
- STUNKARD, A. J. Eating patterns and obesity. *Psychiatric Quarterly*, 1959, 33, 284-295. (a)
- STUNKARD, A. J. Obesity and the denial of hunger. *Psychosomatic Medicine*, 1959, 21, 281-289. (b)
- STUNKARD, A. J. The results of treatment for obesity. *Archives of Internal Medicine*, 1959, 103, 79-85. (c)
- SUCZEK, R. F. The personality of obese women. *American Journal of Clinical Nutrition*, 1957, 5, 197-202.
- THORPE, J. G., SCHMIDT, E., BROWN, P. T., & CASTELL, D. Aversion relief therapy: A new method for general application. *Behavior Research and Therapy*, 1964, 2, 71-82.
- WOLFE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.
- YOUNG, C. M., MOORE, N. S., BERRESFORD, K., EINSET, B. McK., & WALDNER, B. G. The problem of the obese patient. *Journal of the American Dietetic Association*, 1955, 31, 1111-1115.

(Received November 13, 1967)

New and forthcoming titles in clinical and abnormal psychology



PRINCIPLES OF BEHAVIOR MODIFICATION

By Albert Bandura, Stanford University

Dr. Bandura presents a comprehensive analysis of the basic psychological principles governing behavior within a social-learning framework. Principles of social psychology, learning, personality, and clinical psychology are integrated to form a unified theory of behavior change.

July 1969 640 pages \$10.00 (tent.)

FOUNDATIONS OF ABNORMAL PSYCHOLOGY

Edited by Perry London, University of Southern California, and David Rosenhan, Swarthmore College

This multi-authored text, edited and integrated by Professors London and Rosenhan, deals with positive normality as well as traditional psychopathology. The aim of the text is to examine abnormality from the vantage point of what is known in general psychology of personality, human neurophysiology, and genetics.

1968 656 pages \$9.95

METHODS OF MADNESS

The Mental Hospital as a Last Resort

By Benjamin M. Braginsky, Wesleyan University, Dorethea D. Braginsky, Fairfield University, and Kenneth Ring, University of Connecticut

In describing how mental patients behave, both within and outside the hospital, the authors make it irrefutably clear that when attempting to account for even the most mundane activities of mental patients, the dominant psychiatric conception of mental illness is not only inadequate but inappropriate.

April 1969 224 pages \$5.00 paper (tent.)

PSYCHOTHERAPEUTIC AGENTS

New Roles for Nonprofessionals, Parents, and Teachers

Edited with commentaries by Bernard G. Guerney Rutgers, The State University

Through articles and editorial commentary, Professor Guerney explores remedial methods and procedures based on the strategy of using nonprofessionals as trained and supervised agents of the professional psychotherapist.

April 1969 608 pages \$9.95

HELPING AND HUMAN RELATIONS

A Primer for Lay and Professional Helpers
Volume 1: Selection and Training

By Robert Carkhuff, University of Massachusetts, Amherst

The first of a two-volume series, this volume deals creatively and pragmatically with the effective selection and training of lay and professional personnel for social work, counseling, clinical work, psychiatry, as well as other areas of human relations.

August 1969 288 pages \$7.50 (tent.)

CLINICAL INFORMATION PROCESSING BY COMPUTER

An Essay and Selected Readings

By Benjamin Kleinmuntz, Carnegie-Mellon University

Focusing on selected applications of computers in clinical decision-making, the author combines an essay on the use and potential utility of these machines with appropriate readings written by scientists active in the field.

May 1969 416 pages \$5.00 (tent.)

Please ask your HRW representative
for an examination copy or write to College Promotion

Holt, Rinehart and Winston, Inc.

383 Madison Avenue, New York, New York 10017

EUROPE

To utilize the low group fares of the major airlines — especially in connection with the International Congress of Psychology in London July 27, and the Annual American Psychological Association Convention in Washington August 31, 1969 — we have scheduled several groups to London, the Continent, the Iberian Peninsula, Scandinavia, Yugoslavia and Greece. Round trip jet fares start at \$230 (New York/London.) Write to Division "P" Travel Wholesalers International, ITT Building, Washington, D.C. 20036, or call collect (202) 296-9161.

Journal of Abnormal Psychology

Library No. 267
Date... 12-6-71
lib
Bureau of Psych. Res.

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
*New York State Psychiatric Institute
and Columbia University*

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt

Journal of Abnormal Psychology

Department of Psychology

N. Y. State Psychiatric Institute & Columbia University

722 West 168th St.

New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR
Managing Editor

MYRTLE HENDRICKS
Technical Editor

HENRY S. SALMON
Advertising and Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

ILLUSORY CORRELATION AS AN OBSTACLE TO THE USE OF VALID PSYCHODIAGNOSTIC SIGNS¹

LOREN J. CHAPMAN AND JEAN P. CHAPMAN

University of Wisconsin

Practicing psychodiagnosticians ($N = 32$), when surveyed, failed to report observing Wheeler-Rorschach Signs 7 and 8 as accompanying male homosexuality although research evidence indicates that these are valid signs. They instead reported observing Wheeler Signs 4, 5, 16, 19, and 20, which research literature indicates are invalid. These popular invalid signs were found to have much stronger rated, verbal associative connections to male homosexuality than the unpopular valid signs. Six hundred and ninety-three undergraduates (divided among 13 conditions) viewed 30 Rorschach cards on each of which was arbitrarily designated a patient's response and his two symptoms. The Ss "rediscovered" the same invalid Rorschach content signs of homosexuality as the clinicians reported observing in their clinical practice, although these relationships were absent in the experimental materials. They did so regardless of the degree to which the clinically valid signs were valid in the contrived task materials.

The psychodiagnostician who uses psychological tests relies usually on his clinical experience for interpreting test responses. He accumulates observations of the different responses that occur as correlates of various personality characteristics of patients. He later uses this information to infer, at least tentatively, the presence of similar characteristics in patients giving similar responses. To do this, the diagnostician must assume, of course, that he is able to observe and remember which characteristics of test performance occur as correlates of each characteristic of personality. The inference of characteristics of personality from the occurrence of isolated characteristics of performance, instead of interpretation in terms of more complex patterns of test performance, is often termed the "sign" approach. This frequently deprecated, yet widely used, approach to projective test interpretation is the focus of the present study.

A large body of research literature demonstrates that many diagnostic test signs lack

the psychological meanings that clinical observers have claimed for them. It is not surprising, then, that other studies have demonstrated that psychodiagnosticians are usually less able, than they believe, to make valid statements about patients on the basis of tests. Little and Shneidman (1959) made this point very clearly. In an especially well-controlled study, they found that eminent clinicians performed only slightly above chance.

The enormous discrepancy between the reports of clinical observers and the research evidence concerning the meanings of test responses has long been a puzzle. The conflict of evidence has been especially disquieting because clinicians usually show substantial consensus as to specific meanings of various test responses, although these same meanings may have been discredited by research evidence.

The present writers (Chapman & Chapman, 1967) have recently suggested a possible resolution of this enigma by demonstrating a source of massive systematic error in observations of correlations between symptom statements and features of projective test protocols. The projective test studied was the Draw-a-Person Test, which has been largely discredited as a measure of personality by the research evidence (Swensen, 1957). Naive undergraduates viewed a series of 45 Draw-a-Person Test drawings paired with contrived

¹This study was supported by Research Grant MH-07987 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Fred Cosentino for his assistance in the gathering and analysis of the data, and to Richard McFall for his critique of the manuscript.

²Requests for reprints should be directed to Loren J. Chapman, Department of Psychology, University of Wisconsin, Madison, Wisconsin.

TABLE 1
CRITICAL RATIO (z) VALUES FOR WHEELER SIGNS THAT DISTINGUISHED
MALE HOMOSEXUALS FROM HETEROSEXUALS IN THREE STUDIES

Study	z value					
	Sign 7	Sign 8	Sign 10	Sign 17	Sign 19	Sign 20
Wheeler	1.64			1.79		
Davids, Joelson, & McArthur	1.70	2.17	1.65		1.80	
Hooker		1.94				1.90

symptom statements about the alleged patients who drew them. In these contrived materials there was no correlation between the occurrence of any symptom statement and any drawing characteristic. The undergraduate Ss "rediscovered" the same correlations between drawing characteristics and symptoms that a group of clinicians most often reported observing in their clinical practice. These relationships are called "illusory correlations" because the naive Ss reported observing them, although they were actually absent in the experimental materials. Moreover, the popularity of the various illusory correlations corresponded to strength of rated, verbal associative connection between symptom and drawing characteristic. These findings suggest that the popular meanings of many test signs, as reported by clinicians, are illusory correlations based on verbal associative connection of the test sign to the symptom, rather than on valid observations.

Not all tests, however, are as completely lacking in validity as the Draw-a-Person Test. Performance on most psychodiagnostic tests has some correlation with personality characteristics. In light of the research evidence indicating that clinicians are only moderately successful in the interpretation of psychodiagnostic test performance, one must wonder whether clinicians' observations of valid signs, when such signs are present, is impeded by their proclivity to observe, instead, associatively based illusory correlations. This question is the central interest of the present paper.

To investigate this issue, it was first necessary to choose a symptom and a test for which both valid and invalid signs have been reported by clinical observers. Male homosexuality and Rorschach content analysis were

chosen because they appear to fulfill these criteria. Wheeler (1949) offered 20 Rorschach signs of male homosexuality. Clinicians commonly report several of these signs as substantiated by their own clinical experience, but research evidence strongly supports only two of the signs. Three studies by different investigators (Davids, Joelson, & McArthur, 1956; Hooker, 1958; Wheeler, 1949) have reported statistically interpretable evidence on the validity of all 20 Wheeler-Rorschach content signs, and these studies show some agreement. Table 1 presents the z values for those Wheeler signs that distinguished homosexual and heterosexual groups in one or more of the three studies at the .05 level using a one-tailed test. For Wheeler's study, the z values were computed by the present writers from Wheeler's published data, using the formula for the critical ratio between uncorrelated proportions. Hooker obtained her values using the comparable formula for correlated proportions, after matching pairs of Ss by total number of responses. Davids et al. did not indicate which formula they used.

As seen in Table 1, Wheeler Signs 7 and 8 were both found to distinguish homosexual from heterosexual groups in two of the three studies. Wheeler Sign 7 is a response on Card IV of "human or animal—contorted, monstrous, or threatening," and Wheeler Sign 8 is a response on Card V, W or Center D, of a "human, or humanized animal." Wheeler's example of "humans" is "woman dressed as a bat" which indicates that the humans are animalized humans. Signs 10, 17, 19, and 20 were each found to distinguish the groups in one study but were tested and not found valid in the other two studies. A finding by chance alone of a significant difference for one sign out of 20 is not unexpected. Therefore, for

purposes of this study, Signs 7 and 8 were considered the only clinically valid signs. This conclusion tends to be supported by Reitzell (1949) who reported Signs 7, 8, and 16 as the most discriminating of the 20 Wheeler signs. Unfortunately, her data cannot be analyzed statistically because she reported them in terms of number of responses of each type given by homosexual and non-homosexual groups, rather than the number of Ss in each group who gave each sign.

The hypotheses of this study are:

1. The "popularity" of signs among practicing clinicians has little relationship to the objective clinical validity of the signs, as indicated by research evidence.

2. The most popular signs among practicing clinicians are the ones that have the strongest verbal associative connection to male homosexuality.

3. Naive observers, when presented with contrived Rorschach responses arbitrarily paired with statements of symptoms of the patient who gave each response, erroneously report observing that these same associatively based invalid signs occur as correlates of homosexuality.

4. The naive observers report these associatively based illusory correlations even when the materials are contrived so that other (clinically) valid correlations are present.

PART I

Experiment I: Clinical Observations by Practicing Psychodiagnostics

It was first necessary to learn the kinds of Rorschach content that psychodiagnostics believe they have observed most often in the protocols of male homosexuals. A questionnaire was prepared for circulation among practicing clinicians. The questionnaire was anonymous, but it asked the clinician to list his academic degrees and the years he obtained them, as well as the number of years of his psychodiagnostic experience.

Instructions

Although the more traditional clinical use of the Rorschach is the interpretation of Rorschach determinants, many practicing clinicians in recent years have observed that the content of Rorschach responses is also related to the patient's emotional

problems. Some of these observed relationships have been discussed in published reports, while others have not. We wish to ask about your observations concerning the content of Rorschach responses by two kinds of patients.

1. Have you seen the Rorschach protocols of a number of men who have problems concerning homosexual impulses, either overt or covert?

2. What kinds of Rorschach content have you observed to be prominent in the Rorschach protocols of men with problems concerning homosexual impulses?

3. If possible, would you list some examples of the kinds of responses made by such men.

Altogether 76 copies of this questionnaire were sent to clinicians at internship training centers, academic psychology departments, and other leading clinical installations. Each of 11 recipients received several copies with the request that he circulate them to his colleagues.

Results

Questionnaires were returned by 42 clinicians 32 of whom said that they had seen the Rorschach protocols of a number of men with homosexual problems and were willing to list the kinds of Rorschach content they had observed. Of these 32 clinicians, 22 reported having the PhD degree and 10 the master's degree. They reported 2-29-yr. psychodiagnostic experience, with a mean of 9.1.

The clinicians' most frequent responses are listed in Table 2. Only two clinicians listed one of the two valid Wheeler signs. Both of these were Wheeler Sign 7. Most of the clinicians listed one or more of five invalid Wheeler signs. Two of these five signs have,

TABLE 2
THE FIVE WHEELER-RORSCHACH CONTENT SIGNS
OF MALE HOMOSEXUALITY MOST
FREQUENTLY REPORTED BY THE
CLINICIANS

Wheeler sign	Percentage of clinicians reporting the sign
16 Human or animal anal content	44
20 Feminine clothing	38
4 Humans with sex confused	28*
5 Humans with sex uncertain	16*
19 Male or female genitalia	38

* The percentages listed here are for the clinicians' reports of this content regardless of the card on which they observed it. Wheeler limited Signs 4 and 5 to Card III, but few of the clinicians specified a card. Wheeler described sexual confusion on Card I as Sign 2, but did not describe sexual uncertainty on that card despite the conceptual similarity of these two classes of percepts.

as previously discussed, each received support in only one of three studies, and the other three were supported by none of them.

In order to test the hypothesis that the popularity of the invalid signs of homosexuality is based on high-strength, verbal associative connection with homosexuality, ratings of strength of associative connection were obtained. The following rating item, used for "anality," illustrates the format that was used.

The tendency for "homosexuality" to call to mind "rectum" and "buttocks" is

- a. Very strong.
- b. Strong.
- c. Moderate.
- d. Slight.
- e. Very slight.
- f. No tendency at all.

Rating items were used for both of the two unpopular valid signs as well as for the five popular invalid signs that are listed in Table 2. The wordings of the signs were in some cases rephrased, as in the example above, to facilitate understanding by the raters. These wordings are listed in Table 3. Because of their conceptual similarity, "sexual uncertainty" and "sexual confusion" were represented in the ratings by a single item of "part man—part woman." In addition, there was one item each for "food" and "maps," which were used as filler items in the studies that follow.

The eight items concerned with homosexuality were intermixed with 80 other items in the same format, most of which were not concerned with sexual material. The questionnaire was given to a group of 34 undergraduate students who did not participate in the other studies reported in this paper.

The six associative ratings from a to f were assigned values from 6 to 1 and a mean was computed for each item. Table 3 reports the mean rated associative strength between homosexuality and each of the eight categories.

As seen in Table 3, the popular invalid signs have a much stronger associative connection to homosexuality than either the unpopular clinically valid signs or the two categories of filler items. There is no overlap between the mean rated strength of the popular invalid signs and that of the other signs and

TABLE 3
MEAN RATED STRENGTH OF ASSOCIATIVE
CONNECTION BETWEEN HOMOSEXUALITY
AND CONTENT AREAS

Content area	Mean rated strength
Popular invalid signs	
Rectum and buttocks	4.38
Part man—part woman	3.53
Feminine clothing	3.12
Sexual organs	4.47
Unpopular valid signs	
Part animal—part human	1.93
Monsters	1.68
Filler items	
Food	1.09
Maps	1.09

categories of filler items. The rated associative strength to homosexuality of the lowest rated popular invalid sign (feminine clothing) was compared with that of the highest rated unpopular valid sign (part animal—part human). The difference was significant as indicated by a two-tailed t test ($t = 3.52$, $p < .001$). These data suggest that whether or not a clinician reports a given category of percepts as a correlate of homosexuality is determined primarily by the strength of its verbal associative connection to the symptom, rather than by objective reality. This inference is congruent with earlier findings (Chapman & Chapman, 1967) concerning the Draw-a-Person Test.

PART II

Laboratory studies were designed to determine whether naive observers, presented with contrived statements of patients' symptoms and their Rorschach responses, would make the same errors of observation that the clinicians appear to have made in their observational reports. The demonstration in the laboratory of these systematic errors of observation based on verbal associative connection would lend strong additional support to the contention that the clinicians' reports reflect illusory correlation based on associative connection.

Subjects

The Ss in the laboratory studies were tested under 13 conditions which, for convenience of exposition, were divided into three experiments (Experiments II, III, and IV). All Ss were students in an intro-

ductory psychology course. The number of Ss in Conditions 1-13 were 60, 43, 53, 37, 39, 52, 49, 59, 54, 61, 60, 60, and 66, respectively. No S served in more than one condition. All Ss were asked on a questionnaire if they had any familiarity or experience with principles of Rorschach interpretation. The occasional S who indicated any such familiarity was not included in the sample.

Experiment II (Conditions 1-5): Illusory Correlation in the Absence of Valid Relationships

Experiment II was designed to determine if the invalid signs that were found to be popular with the clinicians would also be reported by naive observers when no valid relationship is present between any category of percepts and the symptom of male homosexuality. This study contained five conditions, each of which was designed to explain, on the basis of illusory correlation, the popularity of the report of one of the five popular invalid signs listed in Table 2.

Method

Clinical materials were fabricated to be shown to naive observers. The materials consisted of 30 Rorschach cards, on each of which one percept (or response) was paired with two statements of the emotional problems of the purported patient who was alleged to have given the response. The cards were covered with transparent plastic. Rorschach percepts were indicated by circling an area of the card and pasting on it a typed statement of the verbalization. For example, for one of the 30 Rorschach responses, the center area of Card V—or area D-7 (Beck, 1961)—was circled and labeled "Bugs Bunny." In a corner of the card appeared the statement:

The man who said this

1. has sexual feelings toward other men.
2. feels sad and depressed much of the time.

The 30 percepts were chosen so that six fell into each of five categories, which were:

- A. One popular invalid sign
- B. Wheeler Sign 7: Human or animal—contorted, monstrous, or threatening on Card IV (a clinically valid sign)
- C. Wheeler Sign 8: Humanized animal or an animalized human on Card V (a clinically valid sign)
- D. Geographic features (a filler category)
- E. Food (a filler category)

There were five conditions in Experiment II, which differed only as to which invalid sign was used. In all other respects, both the procedure and materials in all five conditions were identical. The invalid signs for the five conditions were: Condition

1—human or animal with anal content; Condition 2—feminine clothing; Condition 3—humans on Cards I or III with sex confused; Condition 4—humans on Cards I or III with sex uncertain; and Condition 5—male or female genitalia.

The percepts used for each category are listed below. Locations of percepts are indicated using Beck's system. The percepts chosen for the seven Wheeler signs were Wheeler's (1949) examples, and others modeled after his examples.

Wheeler Sign 7. Human or animal, contorted monstrous or threatening. All were on Card IV, W: (a) a horrid beast, (b) Frankenstein, (c) a headless monster about to step on me, (d) man looking back through legs, (e) a giant with shrunken arms, and (f) a deformed man doing a back bend.

Wheeler Sign 8. Humanized animal or animalized human, on Card V. Four were W responses: (a) woman with butterfly wings, (b) an alligator in a fur coat, (c) man dressed like a bat, (d) a pigeon wearing mittens. Two were D-7: (e) a dog wearing clothes, and (f) Bugs Bunny.

Human or animal anal content. (a) Card I, Dd-22, rectum, (b) Card VII (inverted), Dd-25, anus, (c) Card IV, W, anal opening of man bending over, (d) Card VI, W (without D-3), horse's rear end, (e) Card IX, both D-1, woman's buttocks, and (f) Card II, D-4, anal opening.

Feminine clothing. (a) Card III, D-10, woman's high-heeled shoe, (b) Card II, D-2, woman's hat, (c) Card X, D-9, fur stole, (d) Card VIII, D-5, woman's laced corset, (e) Card III, D-7, woman's bra, and (f) Card VI, D-1 (inverted), woman's fur cape.

Sexual confusion. Three of these percepts were on Card I, D-4: (a) it looks like a woman but with broad shoulders, (b) upper part is male and the bottom is female, (c) this has shoulders like a man but breasts and hips like a woman. Three were on Card III, D-9: (d) looks like a man below the waist but like a woman above, (e) a man but with breasts, (f) a woman, here are her breasts, but has the feet of a man.

Sexual uncertainty. Three of these percepts were on Card I, D-4: (a) I guess this is a man or maybe it's a woman, (b) a person can't tell if it's a man or a woman, (c) a human, could be masculine or feminine. Three were on Card III, D-9: (d) a person, might be male or it might be female, (e) could either be a man or a woman, and (f) two people, but I can't tell what sex they are.

Male or female genitalia. (a) Card I, Dd-22, testicles, (b) Card II, D-3, female genital organ, (c) Card III, Dd-26, penis, (d) Card X, D-11, male genitalia, (e) Card VI, D-2, penis, and (f) Card III, D-8, vagina.

Geographic features. (a) Card VII (inverted), D-2, map of North and South America, (b) Card VII, D-3, map of Spain, (c) Card X, D-9, map of California, (d) Card II (inverted), D-2, map of South America, (e) Card X, D-13, New Zealand, and (f) Card VI (inverted), Dd-25, map of California.

Food. (a) Card VII, D-4, a loaf of rye bread split in half, (b) Card IX, D-6, raspberry sherbet,

(c) Card VIII, D-7, jello, (d) Card VI, D-6, asparagus tips, (e) Card X, D-2, scrambled eggs, and (f) Card III, D-1, bowl of fruit.

The two statements of emotional problems or symptoms listed on the cards were drawn from a pool of four such statements. These were:

- 1. He has sexual feelings towards other men.
- 2. He believes other people are plotting against him.
- 3. He feels sad and depressed much of the time.
- 4. He has strong feelings of inferiority.

The statements of symptoms and Rorschach percepts were paired on the 30 cards so that each of the four symptom statements appeared 15 times. Each symptom statement was paired with three of the six percepts from each of the five categories of percepts. Thus there was no relationship between the occurrence of any one of the four symptoms and any one of the five categories of response.

The prediction was that despite this lack of true correlation in the experimental materials, Ss who viewed the series of 30 cards in each of the five conditions would believe that they observed that the symptom statement, "He has sexual feelings toward other men," had appeared more often with the associatively based invalid sign than with any of the other four categories of content.

The testing was done in groups of 30 or fewer. The Ss were given some brief introductory information as to the nature of the Rorschach. They were told that Rorschach responses indicate personality functioning. However, they were given no information about categories of either content or determinants. They were then told:

I am going to show you a series of inkblots, one at a time. On each inkblot you will find a typed statement of what one patient saw on this blot and also what his two chief emotional problems are. Each of these 30 cards represents a different patient. You will see what 30 different patients said they saw on a card. Now let me tell you what I want you to do. Please carefully study

each inkblot and the statement of what the patient said that he saw in it. Also study the statement of the patient's two severe emotional problems. When everyone has looked at all of the cards, I'm going to give you a questionnaire in which I will ask you about the kinds of things seen by patients with each kind of problem.

The cards were circulated in a prearranged pattern so that each S saw each of the 30 cards for 60 sec. After S had seen the 30 cards, he was given a questionnaire which presented four items (one for each of the four symptom statements) in the following format.

Some of the things in the inkblots were seen by men who have the following problem:

He has sexual feelings toward other men.

Did you notice any general kind of thing that was seen most often by men with this problem? Yes — No — If your answer is yes, name that kind of thing, and give one example of that kind of thing.

Kind of thing _____
Example _____

Two forms of the questionnaire were used with different orders of the four items.

Results

Since the responses to the homosexual item were of primary interest in the present paper, the analysis of the results focused on this item. Table 4 shows the correlates reported for homosexuality by Ss in each of the five conditions. As seen there, very few Ss indicated that they could find no relationship between the percepts and the symptoms. In each condition, Ss, as predicted, reported that they observed the hypothesized illusory correlate as accompanying homosexual problems more

TABLE 4
PERCENTAGE OF NAIVE OBSERVERS REPORTING EACH OF FIVE CATEGORIES OF PERCEPT AS MOST OFTEN ACCOMPANYING THE PROBLEM OF HOMOSEXUALITY FOR EACH OF THE FIVE CONDITIONS IN EXPERIMENT II

Correlate reported	Condition				
	1 Anality	2 Feminine clothing	3 Sexual confusion	4 Sexual uncertainty	5 Genitalia
Predicted illusory correlate	58	40	45	24	51
Clinically valid signs					
Human or animal, monstrous	8	14	13	10	3
Part animal—part human	3	14	6	8	0
Filler categories					
Geographic features	12	5	13	16	13
Food	8	16	15	18	13
Other correlates	0	0	2	10	8
No correlate reported	10	12	6	13	13

often than any other category of percept. Chi-square analysis indicated that the distributions of responses among the five categories departed from chance ($p < .01$) for each condition except Condition 4, in which the hypothesized illusory correlate was sexual uncertainty.

The frequent report of the invalid Wheeler signs as illusory correlates of homosexuality cannot be attributed to a higher frequency of report for all symptom statements. The mean percentage report of the five signs for the other three symptoms were as follows: anality, 4%; feminine clothing, 13%; sexual confusion, 9%; sexual uncertainty, 8%; and genitalia, 10%. In none of the five conditions was the illusory correlate of homosexuality the most frequently reported correlate of any one of the other three symptoms.

One may conclude from the data of Table 4 that the observation of popular invalid signs of male homosexuality were reproduced in the laboratory as illusory correlates. On the other hand, the clinically valid signs of homosexuality were not reported as illusory correlates of that symptom any more often than were the two filler categories.

Experiment III (Conditions 6-11): Illusory Correlation in the Presence of Valid Relationships

Experiment III was undertaken to determine if the associatively based illusory correlates are reported by naive observers even if the clinically valid signs have contrived validity in the experimental task materials. To build in contrived validity, the clinically valid Wheeler Signs 7 and 8 were paired with the symptom statement of homosexuality more often than with the other classes of percepts.

Method

Only two of the five popular invalid signs were chosen to be used in Experiment III. They were anality and sexual confusion.

The task materials were almost identical to those used in Experiment II. The instructions and answer sheets were unchanged. The percepts, as well as the pool of symptom statements were the same as those previously used in Conditions 2 and 4 in which anality and sexual confusion were used as popular invalid signs. As in the earlier conditions, each *S* saw 30 cards, six presenting a percept of a popular invalid sign, six for each of the two clinically valid

signs (Wheeler Signs 7 and 8), and six for each of the two filler categories (geography and food). The task materials differed from Experiment II only in the percentage of clinically valid percepts that were accompanied by the symptom statement of homosexuality. Three degrees of contrived validity were investigated: one in which the symptom statement of homosexuality occurred with two-thirds of the percepts of each of the two valid signs, one in which it occurred with five-sixths of them, and one in which it occurred with all of them. After being rounded off to the nearest percentage point the values of the three degrees of contrived validity were 67%, 83%, and 100%, respectively. The combination of three degrees of contrived validity with each of two popular invalid signs yielded six conditions altogether. In Conditions 6, 7, and 8, the popular invalid sign was sexual confusion, with contrived validity for Signs 7 and 8 of 67%, 83%, and 100%, respectively. In Conditions 9, 10, and 11, the popular invalid sign was anality with the same three degrees of contrived validity for Signs 7 and 8. As in Experiment II, 50% of each of the other three categories of percepts were accompanied by the symptom statement of homosexuality, as well as by the other three symptom statements.

Results

Table 5 shows the percentage of *Ss* in each condition who reported the co-occurrence with homosexuality of each of the five categories of percepts. The most striking feature of these data is the degree to which the illusory correlates based on associative connection are impervious to the contrary influence of valid correlations. In all six conditions, the clinically popular invalid sign was the most frequently reported correlate of the symptom statement.

For each of the six conditions, the report of the homosexual symptom statement was distributed among the various categories of percepts significantly different from chance, as shown by chi-square analyses ($p < .01$ in every case). The number of *Ss* who chose the invalid sign as a correlate was compared with the number who chose one or the other of the two valid signs. A goodness-of-fit chi-square was used with the expected values being one-third of the *Ss* for the invalid sign, and two-thirds of the *Ss* for the two valid signs combined. In five of the six comparisons the difference was significant ($p < .01$ in each case). The one nonsignificant difference was for anality when the clinically valid signs had 83% contrived validity. The difference for this condition fell short of significance ($\chi^2 = 3.03$, $df = 1$, $p < .10$). These findings

TABLE 5

PERCENTAGE OF Ss IN EXPERIMENT III REPORTING
THE OBSERVATION OF EACH CATEGORY OF
PERCEPT AS A CORRELATE OF
HOMOSEXUALITY

Category	Percentage contrived validity for Signs 7 and 8		
	67%	83%	100%
Conditions 6, 7, and 8			
Invalid signs			
Sexual confusion	50	41	46
Valid signs			
Human or animal—monstrous	12	27	20
Part animal—part human	10	6	14
Filler categories			
Geographic features	2	12	5
Food	8	0	2
Other correlates	10	6	8
No correlate reported	10	8	5
Conditions 9, 10, and 11			
Invalid sign			
Anality	54	34	55
Valid signs			
Human or animal—monstrous	16	16	15
Part animal—part human	13	23	13
Filler categories			
Geographic features	6	13	5
Food	2	3	2
Other correlates	0	3	2
No correlate reported	9	7	8

indicate that the invalid signs were reported as correlates of homosexuality more often than were the valid signs, which had objectively true correlations with the symptom.

Another way to view these data is in terms of a comparison of the number of Ss who reported the popular invalid signs, and the number who reported the valid signs, as correlates of homosexuality at each of the four levels of contrived validity of the two valid signs. (The four levels are the 50% co-occurrence of the clinically valid sign with the symptom statement—from Experiment II—plus the 67%, 83%, and 100% levels of contrived validity of Experiment III.) When the popular invalid sign was sexual confusion, these four conditions did not differ on frequency of report of the invalid sign ($\chi^2 = .84$, $df = 3$, $.80 < p < .90$). They also did not differ on number of Ss reporting one or the other of the two valid signs ($\chi^2 = 4.92$, $df = 3$, $.10 < p < .20$). When the popular in-

valid sign was anality, the four conditions of contrived validity differed in both the number of Ss reporting the popular invalid sign ($\chi^2 = 8.47$, $df = 3$, $p < .05$) and in the number reporting one or the other of the two valid signs ($\chi^2 = 13.05$, $df = 3$, $p < .01$). Inspection of the data for the three anality conditions indicates that Ss achieved highest accuracy by both criteria in the 83% validity condition. These findings indicate that increasing the contrived validity had a small effect in reducing the report of one popular invalid sign as a correlate of homosexuality, but not the other. The validity was accurately perceived to a slight degree.

Further analyses were performed to seek evidence on the accuracy of perception of the contrived validity. For the three conditions for which sexual confusion was the popular invalid sign, more Ss chose the valid sign than the filler items for the 100% validity condition ($\chi^2 = 9.38$, $df = 1$, $p < .01$) but this difference did not emerge on either the 67% or 83% validity conditions. For the three conditions for which anality was the popular invalid sign, Ss reported the valid signs more often than the filler categories for all three levels of contrived validity ($p < .05$ in all three cases). These findings again show that the contrived validity was to some degree accurately perceived, despite the fact that its presence did little to reduce the frequency of report of the illusory correlates.

Experiment IV: Accuracy of Report in the Absence of Popular Invalid Signs— Condition 12

The most striking findings of Experiment III were the frequency of report of the illusory correlates, and the low accuracy of Ss observational reports of the valid signs. These findings might lead one to wonder if the infrequent detection of valid signs on these tasks is due to the distracting influence of illusory correlates or whether the valid signs are inherently difficult to discover. Further, one might wonder whether associatively based illusory correlation occurs only when the detection of valid signs is very difficult.

In defense of the tasks used in Experiment III, one might point out that they surely present the observer with a less difficult job

of information processing than does conventional clinical practice. In clinical practice, the symptoms are more numerous and more ambiguous, and many percepts are given by each patient. Also, the clinician encounters the patients over a long period of time, so that retrospective falsification of observations should occur more often. Nevertheless, it is of interest to determine if accuracy is higher on a task that is comparable to those of Experiment III, but in which associatively based popular invalid correlates are absent. The first part of Experiment IV (Condition 12) was designed to give evidence on this question.

Method

The stimulus materials were almost identical to those of the two 83% validity conditions (Conditions 7 and 10) of Experiment III. The symptom statement of homosexuality accompanied 83% of the percepts of the two clinically valid signs (Signs 7 and 8) and 50% of the other categories of percepts. The only change from Conditions 7 and 10 was that in Condition 12, nonsexual body parts were substituted for a popular invalid sign. The percepts for the category of nonsexual body parts were: (a) Card II, D-4, hands, (b) Card VII, D-1, gray hair, (c) Card IX, Dd-31, nose, (d) Card I, Dds-30, eyes, (e) Card II, D-2, foot, and (f) Card VI, Dd-25, toe. Thus, each *S* saw five classes of percepts: nonsexual body parts, humanized animals, monsters, food, and geographical features. The mode of administration of the task was identical to that of Experiment III.

Results

Table 6 shows the percentage of *Ss* who reported each category of percept as a correlate of homosexuality. The distribution of *Ss* among the five categories differed from chance ($\chi^2 = 26.23$, $df = 4$, $p < .001$). As seen in Table 6, the two clinically valid signs were the most frequently reported categories, and 65% of the *Ss* reported one or the other of them. This value is almost double the percentage reported in the comparable conditions of Experiment III, and the rise in accuracy was significant ($\chi^2 = 11.6$, $df = 1$, $p < .001$).

Condition 13

The second part of Experiment IV was designed to eliminate one possible source of doubt concerning the meaning of the findings in Condition 12. One might speculate that the

TABLE 6
PERCENTAGE OF *SS* REPORTING EACH CATEGORY OF PERCEPT AS A CORRELATE OF HOMOSEXUALITY IN THE TWO CONDITIONS OF EXPERIMENT IV

Category	Percentage of <i>Ss</i> reporting each category of percept	
	Condition 12 ^a	Condition 13 ^b
Clinically valid signs		
Part animal—part human	27	17
Human or animal—monstrous	38	18
Filler categories		
Body parts	15	23
Food	3	24
Geographic features	8	9
Other correlates	2	4
No correlate reported	7	4

^a Eighty-three percent contrived validity for Signs 7 and 8, 50% for the other three categories.

^b Fifty per cent contrived validity for all five categories.

greater accuracy in Condition 12 than in Conditions 7 and 10 is not attributable to accurate observation of the valid correlates, but to illusory correlation. Table 3 showed that the clinically valid signs have a slightly stronger verbal associative connection to homosexuality than do the two filler categories. This associative connection might produce illusory correlation when stronger verbal associates are absent.

Method

In Condition 13 the percepts were identical to those used in Condition 12, but each class of percept appeared equally often with each symptom statement. If the increased accuracy of report of valid correlates in Condition 12 was an artifact of illusory correlation, it should remain in this condition. If the increased accuracy of report reflected increased accuracy of observation, it should disappear.

Results

The results appear in Table 6. The distribution of people reporting the five categories of percepts did not differ from chance ($\chi^2 = 5.17$, $df = 4$, $p > .05$), and the two clinically valid signs were reported at almost exactly the mean frequencies of the other three categories of correlates. This finding demonstrates that the increased accuracy of report in Condition 12 was due to increased accuracy of observation, and was not attributable to illusory correlation.

DISCUSSION

There was a marked congruence between the reports of the clinicians as to their observations in clinical practice and the reports of the naive observers in the contrived experimental situations. Almost none of the 32 practicing psychodiagnosticians reported observing either of the two signs of male homosexuality that research findings indicate are clinically valid. Instead, they tended to agree with one another in reporting several signs that appear invalid in published research, but which have a high strength verbal associative connection to the symptom of homosexuality.

The naive observers presented with contrived clinical materials reported similar erroneous observations. They reported the associatively based signs as illusory correlates of homosexuality both when there were no valid relationships present and when the two clinically valid signs had a contrived validity in the task materials.

One of the most striking findings of these studies is the persistence of illusory correlation in the face of contradictory reality. Even in the two conditions in which all of the percepts of the two valid signs were paired with homosexuality, these two valid signs were reported as correlates of the symptom less often than the associatively based invalid signs. In addition, the low accuracy of report of the valid signs cannot be attributed entirely to there being too much information for the observer to process. Instead, the low accuracy is, to a considerable degree, a result of Ss' susceptibility to the illusory correlation. Experiment IV showed that, when the percepts that have a high associative connection to the symptom were not presented, the observers almost doubled their accuracy

of report of valid correlates. The illusory correlates blind the observer to the presence of valid correlates of the symptom.

Associatively based illusory correlation is a powerful bias in the observational report of correlations between classes of events. Yet its influence is so unapparent that many practicing psychodiagnosticians have overlooked it, and have substituted illusory correlates for valid correlates in their diagnostic practice. One possible solution to this problem might be to demonstrate to graduate students, as part of their training, their own propensities toward illusory correlation.

REFERENCES

- BECK, S. *Rorschach's test*. Vol. 1. *Basic processes*. (3rd ed.) New York: Grune & Stratton, 1961.
- CHAPMAN, L. J., & CHAPMAN, J. P. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, 72, 193-204.
- DAVIDS, A., JOELSON, M., & MCARTHUR, C. Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics, and normal males. *Journal of Abnormal and Social Psychology*, 1956, 53, 161-172.
- HOOVER, E. Male homosexuality in the Rorschach. *Journal of Projective Techniques*, 1958, 22, 33-54.
- LITTLE, K. B., & SHNEIDMAN, E. S. Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs*, 1959, 73(6, Whole No. 476).
- REITZEL, J. M. A comparative study of hysterics, homosexuals and alcoholics using content analysis of Rorschach responses. *Rorschach Research Exchange*, 1949, 13, 127-141.
- SWENSEN, C. H. Empirical evaluations of human figure drawings. *Psychological Bulletin*, 1957, 54, 431-466.
- WHEELER, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Research Exchange*, 1949, 13, 97-126.

(Received February 29, 1968)

THINKING DISORDERS IN PARENTS OF YOUNG PSYCHOTIC CHILDREN¹

ERIC SCHOPLER AND JULIE LOFTIN

University of North Carolina School of Medicine

In this study parents of young psychotic children were tested on the Object Sorting Test. The sample differed from the replicated studies in that parents of adult schizophrenics were tested in those studies. Since childhood psychosis has substantial differences from adult schizophrenia, it was anticipated that parents of psychotic children would not show the same extent of thought impairment found in parents of adult schizophrenics. Parents of psychotic children, however, showed more thought impairment than parents of normal children. Mothers of psychotic children showed more impairment than fathers. Findings were interpreted to suggest circumscribed test anxiety in association with a psychotic child rather than representing a formal thought disorder.

The interaction between schizophrenics and their parents has received considerable attention during the past 10 yr. These parents have been found to be rejecting and emotionally cold (Kanner, 1943). They involve their children in a "double bind" (Ruesch & Bateson, 1951) in which the child can make no correct response; they engage in pseudomutuality (Wynne & Singer, 1963) where interpersonal communication is garbled; and they have been found to suffer from "parental perplexity" (Goldfarb, 1958), including attitudes of uncertainty, lack of spontaneity, and lack of appropriate empathy and awareness of the child's needs. Bettelheim (1950) has been most articulate about linking parental pathology with the etiology of the extreme stress which produces the schizophrenic development in children. His treatment resolution has been to conduct a parentectomy and to provide residential treatment for the child which prevents his return to the toxic parental influence.

The above literature systematically deletes any consideration of schizophrenia as based on biogenetic impairments. No consideration is given to family members as being reactive

to the schizophrenics' elusive deficiencies. Rimland (1964) has argued that parents of autistic children have been held responsible for being the primary cause of their child's autism, although the origins of the condition were to be found in genetic and organic sources. Frank's (1965) review of the literature also challenged the role of parental pathology in the formation of childhood pathology. He argued that research evidence on parental contribution to childhood disorders was not consistent with the emphasis given by psychodynamic theories and formulations. This etiological emphasis has face validity, especially when considering disordered thinking in parents of young, psychotic children. These parents most often function within normal limits and also have other normal children.

The process of disordered thinking has been considered a central mechanism by which pathology is transmitted from parent to child. Thought disorders which may manifest themselves in illogical concepts, fragmented or blurred attention, inability to interpret and communicate the meaning of objects and events, have been for many investigators the cardinal characteristics of schizophrenia. Lidz, Cornelison, Terry, and Fleck (1958) presented lucid clinical examples of how the parents of schizophrenics maintained self-concepts and thinking styles which were incompatible with reality. As the parents distorted reality to suit their own needs, their children received continuous training in irra-

¹ This project was partially supported by General Research Support Award No. FR-05406 from the United States Public Health Service. The contribution of Irving Gottesman to the formulation and design of this study is gratefully acknowledged.

² Requests for reprints should be sent to Eric Schopler, Department of Psychiatry, University of North Carolina Medical School, Chapel Hill, North Carolina 27514.

tionality. Nine of the 15 patients had at least one parent who was considered schizophrenic, paranoid, or ambulatory schizophrenic. None of these parents, however, had ever been hospitalized in a mental hospital. They all belonged to the middle or upper middle class and were functioning adequately socially. Since the finding of schizophrenogenic thought disorders in the parents was based on sophisticated clinical observation and analysis, it was not at once obvious that explicit and objective measures of thought disorder would support this clinical finding.

Lovibond (1954) developed such an objective scoring system for measuring impairment in conceptual thinking. He defined conceptual thinking in Pavlovian terms. A concept was considered a cortical representation of a generalized relationship between the objects, events, and processes of the external world. He regarded conceptual thinking as the highest or third stage in the process of acquiring subjective representations of the real world. The first stage was simple unanalyzed sensation, followed by perception, in which the formation of complex systems of cortical linkages permitted the registration of objects. At this stage *S* was still bound to immediate experience, until the secondary signaling system, based on language, freed the individual from the particularities of concrete experience.

Lovibond used the Object Sorting Test (OST) as described by Rapaport (1945) for measuring impairments in conceptual thinking. He developed an objective scoring method for rating the type of impairment in *S*'s response. Both administration of the OST and the scoring system were described by Lovibond (1954). In his pilot study, Lovibond tested 32 schizophrenics hospitalized for electroconvulsive therapy (ECT) or insulin treatment and correlated their test scores with psychiatric ratings of thinking disorder. He found agreement between test results and clinical ratings. Both showed significantly higher impairment scores for schizophrenics than for a normal control group. McConaghy (1959) extended the use of the OST to the parents of schizophrenics. He selected 20 schizophrenics who showed clear indications of thought disorder. In testing their parents,

he found they obtained significantly higher impairment scores than did a randomly selected control group. At least one parent of each schizophrenic *S* reached the impairment level. McConaghy suggested that this finding demonstrated a genetic transmission of thought disorder, specifically in the dominant rather than the recessive mode.

The McConaghy experiment was replicated by Lidz, Wild, Schafer, Rosman, and Fleck (1963) with improved controls. Hospitalized schizophrenics who showed disordered thinking clinically and on projective tests were tested with the OST. Parents of nonschizophrenics were tested to control for the effect of parenthood itself, and the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) was given to control for the effects of intelligence. Only 5 of the 20 parents in the experimental group, in comparison with 12 out of 20 in the McConaghy series, had impairment scores. Hence the McConaghy findings were not confirmed. When the protocols were rescored by Lovibond, however, it turned out that he tended to assign higher scores for faulty grouping. When a new set of raters used the corrected Lovibond scoring system, the McConaghy findings were confirmed. His interpretation of a dominant genetic transmission was maintained.

Because of the great potential value of this technique for studying the etiology of psychosis and also for epidemiological investigations, a second replication was conducted by Rosman, Wild, Ricci, Fleck, and Lidz (1964). Using a larger sample than before, they investigated the relationships of sex of parent, age, intelligence, and social class to the test measures of thought disorder. A higher frequency of pathological scores was found among mothers of patients. On the other hand, there was no difference between fathers of patients and control fathers or mothers. The age of the parents did not affect their impairment scores. Their intelligence, however, did. At the lower levels of intelligence, parents in both groups obtained higher impairment scores and could not be distinguished from each other. Only in the higher levels of intelligence, education, and occupation could the patient-parent group be differentiated from the control group. The evidence to date

is clear that a greater incidence of illogical and inappropriate conceptual thinking is found among parents of psychotic patients than among parents of nonpsychotic patients. Rosman et al. (1964) correctly inferred that the findings neither confirm nor deny genetic sources of schizophrenia.

In the previous studies the experimental groups were parents of adult schizophrenics. The present study is a replication, using parents of children diagnosed as psychotic or autistic. There is growing evidence in the literature that early childhood psychosis may not be the antecedent condition to schizophrenia in adulthood. Brain damage (Pollack, 1967) and retardation (Gittelman & Birch, 1967) may play a greater determining role in childhood psychosis. According to Rimland (1964) neurological and biochemical factors are the major determinants of infantile autism. If this is true, it might be expected that parents of psychotic children will not show the same impairment in conceptual thinking shown by parents of adult schizophrenics. Although parents of psychotic children have often been characterized by their rejecting, perplexed, and other pathological child-rearing attitudes, these may be primarily parental stress reactions to an unresponsive child. In such cases the parents' formal thinking processes may be within normal limits even though they have deviant responses to their child.

TABLE 1

AGE, SOCIAL CLASS, YEARS OF EDUCATION, AND IQ OF THE THREE PARENT GROUPS

Experimental groups	Parents' age	Parents' social class	Parents' years of education	Parents' IQ
Parents of psychotics (<i>N</i> = 34)				
Range	26-51	1-4	10-21	79-134
<i>M</i>	36	2	14	111
Parents of retardates (<i>N</i> = 22)				
Range	24-49	1-5	8-21	80-125
<i>M</i>	36	3	14	110
Parents of normals (<i>N</i> = 42)				
Range	24-48	1-3	12-25	102-131
<i>M</i>	34	1	18	121

This proposition was assessed by measuring impairment of conceptual thinking in parents of psychotic children, parents of retarded children, and parents of normal children, using the OST. If this proposition is true, it could be anticipated that these three groups of parents would not differ significantly in their manifestation of disordered thinking. The frequency with which at least one parent of a schizophrenic showed impaired thinking in the McConaghy study (1959) should not be applicable to the parents of young psychotic children. Where the mothers showed greater thinking impairment than the fathers (Rosman et al., 1964), such sex differences were not expected in the present sample.

METHOD

Subjects. The experimental group consisted of parents of 17 children who had been diagnosed as autistic or psychotic. These children had been in a clinic program for long-term group therapy (Speers & Lansing, 1965) for a period from 3 to 5 yr. Although most of these children had improved during their course of therapy, they had all shown at least four of the nine points (Creak, 1961) defining childhood psychosis at some time during their treatment. Their parents had been involved in conjoint group therapy. The families were primarily in the middle class, with above average intelligence, and had other nonpsychotic children in the family.

Two control groups were used. The first included parents of 11 retarded children. This group was selected to be comparable to the experimental group with respect to the age of the impaired child, and the social class status (Hollingshead, 1957) of the parents. The second control group was selected from parents of children enrolled in the public schools or kindergarten, who were of comparable age to those in the experimental group. All children were under 12 yr. of age, with the mean age for each group at 6 yr., 6 mo. The parents of 21 children volunteered to participate. The characteristics of the three groups are summarized in Table 1.

Procedures and measures. The OST materials and method of administration were the same as described by Lovibond (1954), McConaghy (1959), and Rosman et al. (1964). The *S* is presented with an array of 33 common subjects such as a bicycle bell, a hammer, and a screwdriver. In Part I of the test, *E* selects 1 object in a standard sequence and tells *S* to place with it the other objects which belong with it. After each sorting *S* is asked, "Why do these belong together?" In Part II of the test, *E* presents *S* with different groups of objects, each group sharing a common characteristic such as redness or roundness. For each grouping *S* is asked,

TABLE 2
COMPARISON OF OBJECT SORTING TEST SCORES FOR THE THREE PARENT GROUPS

Experimental groups	Impairment scores ^a		Median split ^b	
	0-6	7 or higher	<Mdn	≥Mdn
A. Parents of psychotics (<i>N</i> = 34)	20 (59%)	14 (41%)	10 (29%)	24 (71%)
B. Parents of retardates (<i>N</i> = 22)	15 (68%)	7 (32%)	9 (41%)	13 (59%)
C. Parents of normals (<i>N</i> = 42)	37 (88%)	5 (12%)	23 (55%)	19 (45%)

^a Comparison between Groups A and C: $\chi^2 = 8.59$, $p < .01$.

^b Comparison between Groups A and C: $\chi^2 = 4.91$, $p < .05$; *Mdn* = 3.

"Why do all of these belong together?" and his responses are recorded verbatim.

The Shipley-Hartford Intelligence test (Shipley & Burlingame, 1941) was also administered to all Ss. This test correlates well with other intelligence tests, and provides a quick and reliable estimate of Ss' intelligence.

Test scoring. The scoring system for the OST has been described in detail by Lovibond (1954). Scores are given for eight types of impaired responses which display concept formation based on inappropriate, irrelevant, and illogical relationships among objects. These response types are categorized as syncretistic, fabulatory, chain, symbolic, arbitrary, personal association, irrelevant, and impersonal association. Responses are given weighted scores from 1 to 3, depending on the extent of the conceptual impairment. A S's total score is the sum of all his impairment scores. Correct responses and failures to give a response receive no score. Lovibond (1954) determined empirically that a score of 7 or higher correlated with the clinical rating of disordered thinking in schizophrenic patients and distinguished them from normal controls. The criterion of a score of 7 or higher as an impairment indicator was used by McConaghy (1959), Rosman et al. (1964), and the present study.

All protocols were given code numbers and scored independently by a second *E* who did not know which tests were those of parents of psychotic children, parents of retarded children, or parents of normal children. A Pearson correlation between the scores of both *Es* was calculated and a scoring reliability of $r = .93$ was achieved.

RESULTS

The three groups were compared for both the frequency and magnitude of their impairment scores. Lovibond (1954) used an empirically determined cutoff score of greater than 7. In addition to this cutting score, a median split is reported. Either high or low scores could be expected to be a random distance from the three groups' median (*Mdn* = 3). The magnitude of this distance offers a more sensitive measure of relative impairment

than the cutting score. The impairment scores of each group are presented in Table 2.

Among the parents of psychotics, 41% attained an impairment score greater than 7 while among the parents of normals 12% of the individuals scored higher than 7. This difference was significant at the .01 level. Comparing the same two groups, 71% of the psychotics' parents scored equal to or above the median, while only 45% of the normals' parents attained scores of this magnitude ($\chi^2 = 4.91$, $p < .05$). (All χ^2 analyses were made using two-tailed test). None of the other comparisons between parent groups showed any significant difference for either of the two analyses. Since these findings were consistent with those reported by Rosman et al. (1964) and McConaghy (1959), they did not offer a basis for differentiating parents of adult schizophrenics from parents of young psychotic children.

Table 3 presents the frequency of parent couples in each of the three groups with at least one high-scoring member and at least one member with a score greater than the median.

At least one parent scoring 7 or more was found in 59% of the parents of psychotics, while among the parents of normals, 24% had at least one high-scoring parent ($\chi^2 = 4.81$, $p < .05$). A parallel difference was found for the same two groups for scoring higher than the median ($\chi^2 = 6.66$, $p < .01$). There was no significant difference in any other group comparisons in Table 3. Clearly, the expected similarities among the three groups did not obtain. At least one parent showed thought impairment more often in the psychotics' parents than in the control couples, similar to

TABLE 3
FREQUENCY OF AT LEAST ONE HIGH-SCORING MEMBER AMONG COUPLES
OF THE THREE PARENT GROUPS

Experimental groups	Impairment scores ^a		Median split ^b	
	0-6	7 or higher	< <i>Mdn</i>	≥ <i>Mdn</i>
A. Parents of psychotics (<i>N</i> = 17)	7 (41%)	10 (59%)	5 (29%)	12 (71%)
B. Parents of retardates (<i>N</i> = 11)	5 (45%)	6 (55%)	4 (36%)	7 (64%)
C. Parents of normals (<i>N</i> = 21)	16 (76%)	5 (24%)	15 (71%)	6 (29%)

^a Comparison between Groups A and C: $\chi^2 = 4.81$, $p < .05$.

^b Comparison between Groups A and C: $\chi^2 = 6.66$, $p < .01$; *Mdn* = 3.

the findings reported by McConaghy (1959) and Rosman et al. (1964).

The relationship between sex of parents and thought impairment is shown in Table 4. This table shows comparative findings for mothers and fathers in the three groups.

Fifty-three percent of mothers of psychotics had an impairment score of more than 7, while 19% of mothers of normals and 5% of fathers of normals achieved such a score. These are both significant differences. In the same comparison for scores higher than the median, the mothers of psychotics were higher ($p < .001$). Twenty-nine percent of the psychotics' fathers scored 7 or higher, while 19% of normals' mothers and only 5% of the normals' fathers had such a high score. The difference between the fathers of psychotics and fathers of normals was significant at the .05 level, while there was no difference between fathers of psychotics and mothers of normals. In the comparison of scores greater than the median, there were no differences among fathers of psychotics, fathers of normals, and

mothers of normals. Clearly, there is more thinking impairment in the mothers of psychotics than in the fathers of psychotics. This conclusion is further supported by comparing mothers and fathers of psychotics. Although these mothers have more impairment scores than their husbands, this difference was not significant. However, the mothers of psychotics scored higher than the median in 94% of the cases, while their husbands scored above the median in only 47% of the cases ($\chi^2 = 9.06$, $p < .01$). The mothers of psychotics also had higher scores than mothers of retardates. These findings show more indication of thought impairment for mothers of psychotics than any other parent group. This is consistent with the findings of the Rosman group (1964). The mothers of psychotic children show more thought impairment than fathers whether they are mothers of adult schizophrenics or mothers of young psychotic children.

Six out of eight significant differences reported in Table 4 involved the high-scoring

TABLE 4
COMPARISON OF MOTHERS' AND FATHERS' OBJECT SORTING TEST SCORES
FOR THE THREE GROUPS

Experimental groups	Impairment scores ^a		Median split ^b	
	0-6	7 or higher	< <i>Mdn</i>	≥ <i>Mdn</i>
A _m . Psychotics' mothers (<i>N</i> = 17)	8 (47%)	9 (53%)	1 (6%)	16 (94%)
A _f . Psychotics' fathers (<i>N</i> = 17)	12 (71%)	5 (29%)	9 (53%)	8 (47%)
B _m . Retardates' mothers (<i>N</i> = 11)	9 (82%)	2 (18%)	6 (54%)	5 (46%)
B _f . Retardates' fathers (<i>N</i> = 11)	6 (54%)	5 (46%)	3 (27%)	8 (73%)
C _m . Normals' mothers (<i>N</i> = 21)	17 (81%)	4 (19%)	12 (57%)	9 (43%)
C _f . Normals' fathers (<i>N</i> = 21)	20 (95%)	1 (5%)	11 (52%)	10 (48%)

^a Significant comparisons: A_m and C_m, $\chi^2 = 4.78$, $p < .05$; A_f and C_f, $\chi^2 = 4.31$, $p < .05$; B_f and C_f, $\chi^2 = 7.86$, $p < .01$.

^b Significant comparisons: A_m and A_f, $\chi^2 = 9.06$, $p < .01$; A_m and B_m, $\chi^2 = 8.43$, $p < .01$; A_m and C_m, $\chi^2 = 10.99$, $p < .001$ (*Mdn* = 3).

mothers of psychotics, while one involved the fathers of psychotics. It is of interest to note that the only other significant difference was between the fathers of retardates and the fathers of normals. Forty-six percent of the retardates' fathers obtained an impairment score of greater than 7, while only 5% of the normals' fathers did so ($\chi^2 = 7.86$, $p < .01$). No such evidence of impairment was found for mothers of retardates.

No significant correlations were found between impairment scores, IQ, and social class. This was not consistent with the findings reported by the Rosman et al. (1964) study in which the parents of psychotics and their controls could only be distinguished at the higher intellectual and educational levels and not at the lower levels. The absence of confirmation for the Rosman findings in the present study may well be accounted for by the relative homogeneity in higher social class and intelligence of all three groups in this study.

DISCUSSION

This study was begun with the assumption that early childhood psychosis is not primarily the precursor to the schizophrenia of adulthood, and that the thought impairment reported for parents of adult schizophrenics would not show in parents of young psychotic children. Instead, the authors found that the parents of psychotics showed greater signs of impaired thinking than did the parents of normals. There were more couples with at least one high-scoring parent in the group of parents of psychotics than in the other two groups. The impaired thinking was more prevalent for the mothers of psychotics than for any other same-sexed parent group: The greater degree of thought disorder in mothers of psychotics than in the fathers may reflect their greater exposure and involvement with their impaired children. This explanation, however, should also apply to the parents of retardates in which the opposite tendency was found. The fathers showed a greater tendency to high-impairment signs than did mothers. These findings may be explained by the main differences characterizing childhood psychosis and mental retardation. With psychosis, the child does not relate appropriately, involving impairment in both affect and

thinking, while with retardation the child often relates quite adequately but is incapable of normal coping and school achievement. According to Parson's sex-role definition, the mother's role is more involved in expressive and emotional adaptations, while the father's role is more oriented toward task achievement. The psychotic child's relationship impairment may have a greater impact on the mother, while the failure in task achievement may have a more adverse effect on the father-child relationship. Perhaps such adverse responses to a child's impairment can affect the parents' conceptual thinking. A study testing this possibility is described further in the following paragraphs.

The major issue in this study was that the three hypotheses were not confirmed. Instead, the findings were in substantial agreement with those of previous studies based on the parents of adult schizophrenics. Do these results suggest that no significant difference exists between parents of early childhood psychotics and parents of adult schizophrenics, that the two disorders are closely linked with parental thought impairment? Although this possibility is open, another explanation for the findings must be considered first. The parents of the psychotics were in regular attendance with their children in the clinic in which the study was conducted. A similar psychiatric involvement was true for the experimental groups reported in the previous studies. Perhaps the experimental groups' clinical involvement confounded the parents' concern with being evaluated and their guilt for having produced a sick child—in short, test anxiety may have been confounded with a formal thought disorder to be found in all areas of the Ss' thinking. Whether these impairment scores reflect an impairment circumscribed to the context of a psychiatrically ill child or whether they are an attribute of the Ss' formal thinking process in any situation was tested in a subsequent study.

Such a study was initiated. Another group of parents of psychotic children was tested on the OST. Instead of conducting the study within the context of the psychotic child's psychiatric treatment program, this second group of parents was interviewed about one of their normal children. The purpose of the

interview was to elicit parental characteristics associated with successful child rearing before the OST was administered. Less thought impairment was hypothesized for the parents in this second group than was found in the first (Schopler & Loftin, 1969).

REFERENCES

- BETTELHEIM, B. *Love is not enough*. Glencoe, Ill.: Free Press, 1950.
- CREAK, M. Schizophrenic syndrome in childhood. *British Medical Journal*, 1961, 5, 889-890.
- FRANK, G. H. The role of the family in the development of psychopathology. *Psychological Bulletin*, 1965, 64, 191-205.
- GITTELMAN, M., & BIRCH, H. G. Childhood schizophrenia: Intellect, neurologic status, perinatal risk, prognosis, and family pathology. *Archives of General Psychiatry*, 1967, 17, 16-25.
- GOLDFARB, W. Parental perplexity and childhood confusion. In A. H. Esmond (Ed.), *New frontiers in child guidance*. New York: International University Press, 1958.
- HOLLINGSHEAD, A. *Two factor index of social position*. Privately published, New Haven, Connecticut, 1957.
- KANNER, L. Autistic disturbances of affective content. *Nervous Child*, 1943, 2, 217-250.
- LIDZ, T., CORNELISON, A., TERRY, D., & FLECK, S. Intrafamilial environment of the schizophrenic patient: VI. The transmission of irrationality. *Archives of Neurology and Psychiatry*, 1958, 79, 305-316.
- LIDZ, T., WILD, C., SCHAFER, S., ROSMAN, B., & FLECK, S. Thought disorders in the parents of schizophrenic patients: A study utilizing the Object Sorting Test. *Journal of Psychiatric Research*, 1963, 1, 193-200.
- LOVIBOND, S. H. The Object Sorting Test and conceptual thinking in schizophrenia. *Australian Journal of Psychiatry*, 1954, 5, 52-70.
- MCCONAGHY, N. The use of an object sorting test in elucidating the hereditary factor in schizophrenia. *Journal of Neurology and Psychiatry*, 1959, 22, 243-246.
- POLLACK, M. Mental subnormality and "childhood schizophrenia." *Psychopathology of mental development*. New York: Grune & Stratton, 1967.
- RAPAPORT, D. *Diagnostic psychological testing*. Chicago: Year Book Publishers, 1945.
- RIMLAND, B. *Infantile autism: The syndrome and its implication for a neural theory of behavior*. New York: Appleton-Century-Crofts, 1964.
- ROSMAN, B., WILD, C., RICCI, J., FLECK, S., & LIDZ, T. Thought disorders in the parents of schizophrenic patients: A further study utilizing the Object Sorting Test. *Journal of Psychiatric Research*, 1964, 2, 211-221.
- RUESCH, J., & BATESON, G. *Communication: The social matrix of psychiatry*. New York: Norton, 1951.
- SCHOPLER, E., & LOFTIN, J. Thought disorders in parents of psychotic children: A function of test anxiety. *Archives of General Psychiatry*, 1969, 20, 174-181.
- SIMPLEY, W., & BURLINGAME, C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *American Journal of Psychiatry*, 1941, 71, 1313-1325.
- SPEERS, R. W., & LANSING, C. *Group therapy in childhood psychosis*. Chapel Hill: University of North Carolina Press, 1965.
- WYNNE, L., & SINGER, M. Thought disorder and family relations of schizophrenics. *Archives of General Psychiatry*, 1963, 9, 199-206.

(Early publication received July 26, 1968)

PERCEPTUAL SELECTIVITY IN HALLUCINATORY SCHIZOPHRENICS¹

RONALD H. FORGUS²

AND

ALAN S. DEWOLFE

Lake Forest College

*Veterans' Administration Hospital
Downey, Illinois*

A manual for categorizing themes present in hallucinations was constructed using a representative 10 males and 7 females from 49 actively hallucinating patients interviewed. The authors agreed on the major theme for 30 of 32 remaining Ss, indicating significant interjudge reliability ($p < .001$, Cohen's $k = .93$). An hypothesis based on cognitive theory and sensory deprivation research that an hallucinatory theme is congruent with selective recall was supported. Twenty-seven available Ss heard a story with themes from the hallucinatory theme classification manual. In 18 of the 23 Ss on whose dominant recall theme two new judges agreed, the major hallucinatory theme category predicted the dominant recall theme. This significant relationship (Cohen's $k = .74$, $p < .001$) appeared meaningful, rather than spurious, based on internal and external criterion measures.

In clinical work with psychotic patients there is little doubt that some patients do, in fact, hallucinate. Further, the principle of psychic determinism would indicate that this phenomenon which, in most formulations, is analogous to dreams, should be a source of potentially meaningful information which could aid in understanding the patient's personality dynamics. There has, however, been little systematic scientific research on hallucinations. Consequently, most theories about hallucinations have been developed from post hoc interpretation of clinical data in the framework of psychoanalysis (e.g., Arieti, 1955, 1959).

Hebb (1949), a pioneer in the study of cognition and brain functioning, has more recently (Hebb, 1960) issued a challenge to behavioral scientists. More specifically, he suggested that contemporary research in sensory deprivation has brought the scientific study of phenomena such as hallucinations within methodological reach. The approach which Hebb suggests is based in part on the fact that some of the Ss who have been subjected to reduced sensory variation have re-

ported experiences which resemble hallucinations. For a more complete review of these studies see Zuckerman and Cohen (1964). While the exact relationship of these sensory phenomena to clinical hallucinations has not been established, there seems to be general agreement that the human brain needs varied sensory stimulation for organized, logical, goal-directed thinking. If the sensory variation is not sufficient, a confusion between external and internal sources of stimulation can result. An hallucination is a specific instance of this kind of confusion in which the individual attributes an external source for stimulation which, in fact, originated internally.

In sensory deprivation studies, *E* reduced the sensory variation by making the environment more homogeneous, whereas, it is here postulated that the sensory variation of the hallucinatory schizophrenic is reduced by some form of central inhibition of perceptual processes. Recent formulations of brain and perceptual functioning (e.g., Miller, Galanter, & Pribram, 1960; Pribram, 1963) state that only stimuli for which cognitive models are available can be perceived; that is, there is selective gating of environmental information.

It is further postulated that by comparison with normal individuals, hallucinating schizophrenics have limited cognitive models. Therefore, the hallucinating schizophrenic gates

¹ The authors wish to express their appreciation to Erin Kenefick and Roger Kirman, who served as judges, and to Toni Ach Lowenthal, Susan Spilman, and William Valko for their assistance in gathering data.

² Requests for reprints should be sent to Ronald H. Forgas, Lake Forest College, Lake Forest, Illinois.

very little information. The hallucination occurs in the context of the patient's reduction in sensory variation which causes confusion between external and internal sources of information. Since, in the hallucination, an external source is attributed to stimulation which originates internally, it follows logically that hallucinatory themes must be a valid index of the patient's cognitive models.

In view of the foregoing theoretical orientation the present study was undertaken for two reasons: (a) to develop a reliable scoring system for categorizing the dominant theme present in the hallucinations of schizophrenic Ss, as they describe them, and (b) to conduct an initial test of the hypothesis that the accentuation in perceptual selectivity will be congruent with the patient's hallucinatory themes expressed in these categories.

METHOD

Selection of Subjects

Staff psychologists in a psychiatric hospital submitted a list of patients who, in their judgment, met two criteria: (a) were currently hallucinating and (b) had been hallucinating for at least 3 mo. A total of 32 males and 22 females was obtained.

The first author interviewed these Ss individually with the primary purpose of soliciting from each patient a description of his hallucinations. The tape-recorded interviews never exceeded 45 min. The hallucinatory themes were determined for 29 males and 20 females (3 males and 2 females were eliminated because they denied the existence of hallucinations or because they were incoherent).

Procedure

Reliability study of the scoring manual for hallucinations. The taped interviews of the 49 Ss were transcribed. One-third of these cases (10 males and 7 females), representative with respect to social history and distribution of hallucinatory themes, were selected to construct a scoring manual. A shortened version of the manual is presented in Table 1.

Both investigators used the manual to score independently the *major theme* present in the hallucination described in the interview for the 19 males and 13 females remaining. A summary of the social history of these 32 Ss is available from the authors.

Preliminary predictive validity study. Following the establishment of the interrater reliability of the scoring, predictions were made about the direction of selectivity and distortion in the perception of the patients on whose hallucinatory theme categories the two authors had agreed.

To test these predictions, a male and a female story were constructed containing themes matching

TABLE 1

SAMPLES OF SCORING CATEGORIES FOR THEMES OF HALLUCINATIONS

Conscience disturbance (male 6%, female 67%)

Examples would be voices say: "you are a tramp," "you have cancer" (implies you are rotten inside), "you are evil," or "repent your sins," etc.

Compensatory grandiosity (male 17%, female 0%)

Ego ideal (grandiose delusions, etc.). Examples: "you are superman, God, etc.," "you are the best person on earth," "you will live forever," "no one can hurt you," or "you will win, no one can stop you," etc.

Level of maturity: Coping with problems (male 56%, female 17%)

Examples for women: patient hears children's voices crying to her for help, or hears children begging her for food, etc.

Examples for men: voices say "stick up for your rights," hears voices say "find a job and get married," etc.

Impulse control problems (male 17%, female 0%)

Examples: voices verbalize to eat, drink, and be merry (or make Harry or Mary). Voices say "go out with no clothes on," "be licivious," "take a shit," etc.

Denial of frustrated needs, e.g., dependent needs (male 0%, female 0%)

Regressive fantasy content. Examples: voices say "you are nice and sweet and the doctor and nurses are nice and sweet," "your mother will take you home soon," etc.

Conflicting preoccupations (male 0%, female 0%)

Example for women: "be licivious—you are evil."
Example for men: "you are superman—grow up."

Impersonal hallucinations (male 6%, female 17%)

Patient does not see the hallucination as related to him. Patient hears "voices of a crowd cheering," "cars being driven noisily," or patient sees "just clouds and spaces," etc.

Note:—The percentages shown indicate the proportion of males and females whose hallucinatory themes were in that category. $N = 18$ males and 12 females.

six of the hallucinatory theme categories in the scoring manual (the impersonal hallucinatory themes had to be omitted since they could not be made part of the story's content). Each story was tape-recorded by both a male and a female who had no other connection with the current study.

The patients were brought in individually for a second time. The males heard the male story and the females heard the female story. The examiners were two students, a male and a female, who divided the cases equally. The sex of the voice presenting the story and the sex of the examiner were counter-balanced in a Latin-square design. The S was then asked to retell the story immediately after he had

heard it. His reproduction was tape-recorded. The male and female stories are also available from the authors.

The prediction was that the stories would be distorted, the theme remembered, or emphasized being congruent with the dominant theme of the patient's hallucination. Only 25 cases (out of the 30 on which the two investigators agreed) were used for the validity study since 3 cases (two female and one male) had impersonal hallucinations and two females had been discharged before the data could be collected.

The 25 auditory reproductions were transcribed and two new students rated them independently for the dominant theme in the recall. Following the establishment of the interrater reliability of the auditory reproductions, the strength of the relationship between the major theme of the hallucination of each patient and the obtained emphasis in reproduction of his story was assessed.

To guard against the possibility that the distortions in the recall of the hallucinating Ss were spuriously related to hallucinatory content, the story recalls were also elicited from the control group of 25 (17 males and 8 females) nonhallucinating psychiatric patients. A comparison of the dominant recall themes of these two groups was made.

RESULTS AND DISCUSSION

Reliability Study of the Scoring Manual

The scoring manual defined the seven hallucinatory theme categories. The two authors, both clinical psychologists, who independently rated the patients' hallucination descriptions for the main theme present, agreed on 30 of the 32 cases, yielding a binomial z of 13 ($p < .001$).

To express this finding as a reliability coefficient, Cohen's (1960) Kappa (k) coefficient of agreement for nominal scales (a non-parametric statistic) was used. The k value was .93 indicating that, with this sample of Ss, the scoring based on the manual showed high interrater reliability.

Preliminary Predictive Validity Study

The two students agreed on the dominant themes present in the reproduced auditory story for 23 of the 25 cases. Again, the binomial test was computed and the z value was 9.6 ($p < .001$). The k coefficient was .90. Hence, scoring of the auditory reproduction data also had good interrater reliability.

Finally, as a preliminary predictive validity measure, the agreement between the predictions from the dominant theme of the patient's hallucinations and the dominant theme

of his story recall was assessed for the 23 cases on whose recall theme the students agreed. For 18 of the 23 Ss, the predicted and actual dominant recall themes agreed, yielding a binomial z of 7.6 ($p < .001$) and a k coefficient of .74.

Predictions, based on cognitive theory and the empirical results of research on sensory deprivation, were supported in the results of the present study. The possibility remained that the obtained relationship between the patients' hallucinatory themes and their selectivity in recall was spurious.

Support for the interpretation that the relationship between hallucinations and perceptual selectivity was meaningful came from two sources. First, if the congruence were the result of an artifact, then most of the successful predictions and a very small portion of the unsuccessful ones should be found in the high-frequency hallucinatory themes (i.e., immaturity for males and conscience for females), with the converse holding for the infrequent themes. In the data from males, there was little indication of a disproportionately high success to failure ratio in the immaturity category, since 75% of the successes and 67% of the failures came from this dominant hallucinatory theme category. In addition, there were no unsuccessful predictions in the most infrequently used hallucinatory theme categories (i.e., compensatory grandiosity and impulse control). For the females, 50% of the prediction successes came from the conscience category but so did the only prediction failure. Thus, for both sexes, there seemed little indication of the disproportionate success to failure ratio that would be expected if the relationships between hallucinatory themes and selective recall were due to an artifact.

Table 2 indicates that the dominant story recall themes of the hallucinating patients differed significantly from those of control group nonhallucinating patients. When the males of these two groups were compared on the basis of dominant recall theme of immaturity versus all other, the χ^2 of 6.10, $df = 1$, $p < .02$ indicated that the groups differed significantly in dominant recall themes. When the female hallucinating and nonhallucinating patients were compared with con-

science related dominant recall themes versus all other themes, a Fisher's exact test was used because the small number of cases involved resulted in expected cell frequencies so small that the use of the χ^2 statistic was not appropriate. The Fisher's exact p was .05 indicating that for the females as well as for the males, the dominant recall themes of the hallucinating and nonhallucinating patients differed significantly. When the probabilities for males and females were combined using the Lindquist (1940) method, the combined probabilities χ^2 was 23.41, $df = 4$, $p < .001$.

From the standpoint of both internal (prediction of the success to failure ratio within the hallucinatory theme categories) and external (comparison of the dominant story recall themes of the hallucinatory and nonhallucinatory patients) criteria, there was support for a meaningful, rather than a spurious, congruence between the dominant hallucination theme and the dominant story recall theme.

The finding that the dominant theme in the hallucinations of males was different from the one for females seemed noteworthy. Arrested maturity themes predominated in the males' hallucinations, while conscience related themes were most frequent for females. The preoccupation with what they have not done in the hallucinations of the males and the females heightened concern in their hallucinations with what they have "done" parallels findings that the behavior pathology of male schizophrenics tends to center around excessive withdrawal and passivity, while that of the females usually is related to overly aggressive and dominant behavior (e.g., Cheek, 1964; Gross, 1959; Lorr, O'Connor, & Stafford, 1960). It might also be noted that the specific dominant themes in the hallucinations of the males and females center about their apparent inability to conform to the culturally defined appropriate sex roles.

If the theoretical model used to predict the perceptual selectivity of the hallucinatory schizophrenic patients in the present study gains additional empirical support, it would have many potential implications for the interpretation of schizophrenia. Viewing schizophrenic behavior as resulting primarily from a deficiency in cognitive models (or schemas) for coding incoming stimuli, it would follow

TABLE 2
DOMINANT OR ACCENTED THEMES IN THE STORY RECALL
OF HALLUCINATING AND NONHALLUCINATING
PSYCHIATRIC PATIENTS

Patients	Males		Females	
	Immaturity themes	All other themes	Conscience related themes	All other themes
Hallucinating	14	3	6	3
Nonhallucinating	7	10	2	6

Note.—For males $\chi^2 = 6.10$, $df = 1$, $p < .02$; for females Fisher's exact $p = .05$.

that schizophrenic patients lack models for many, and in some cases, most of the stimuli they encounter.³ Thus, they cannot use much of the information potentially available to them, with resultant reduction of sensory variation below the levels needed for distinguishing between external and internal sources of stimuli and for maintaining organized, goal-directed perception and thinking. Using the above orientation, the poor judgment, the inadequate contact with reality, and the primary process thinking, all of which are diagnostic of schizophrenia, would be explained on the basis of a deficiency in cognitive models.

In the current situation, as is the case in almost every situation, no single theoretical system can claim exclusive rights to a particular prediction. However, the cognitive-theory-based prediction of the congruence between the dominant theme of a patient's hallucination and the dominant theme of his story recall, would not seem to follow as directly from current perceptual and personality theories, or from the projective test "theory," that is, the basic assumptions underlying projective testing.

Current perceptual, personality, or projective theory might be extended to cover hallucinations and psychotic patients and to predict possible relationships between themes in patients' hallucinations and selectivity in their perception and recall of stories. It is not enough to talk simply about projection. It is

³ This formulation does not imply that the cognitive deficit is necessarily permanent, nor does it have any direct bearing on the issue of whether either the disease process or the cognitive deficit is cause or effect.

more powerful to predict the direction of selectivity in projection. Certainly, psychoanalysis would predict distortion toward a threat to the ego whereas the authors' theory predicts projection in terms of ego relevance—that is, accentuation would be congruent with existing cognitive models. Theories which emphasize defense rather than vigilance would, in many cases, make just the opposite prediction.

The cognitive theory based on sensory deprivation research used in the present study not only makes a relatively clear cut prediction of congruence but also offers an empirically based explanation for the clinical phenomenon of hallucinations. The cognitive theory proposed here is certainly not the only one that could predict the empirically demonstrated congruence between hallucination themes and selectivity in perception and recall of a story, and it is not the only theory which explains hallucinations. This theory has in its favor, however, that it is (a) based on the findings of relatively rigorous research, (b) both makes the congruence prediction and explains hallucinations and other schizophrenic symptoms reasonably simply and directly, and (c) offers the possibility of considerable heuristic value as a potential basis of additional research in schizophrenia—for example, making specific directional predictions about the nature of projections and the clinical value of the meaning of hallucinations.

The present study demonstrated that it is feasible to use methods from experimental research in studies of clinical phenomena. Additional work will be needed to firmly establish the validity of the specific findings and the assumptions used in making the prediction of a congruence between hallucination themes and selectivity in perception and recall. Following further empirical support, the particular methodology and the guiding theoretical orientations of this study,

which emphasize the central role of cognitive selectivity, might be extended to shed light on the relationship between perception, affect, and motivation in other forms of adaptive and maladaptive behavior in normal and in other abnormal groups.

REFERENCES

- ALDER, G. W., LANE, E. A., CORCORAN, C. J., WERNERT, A. Childhood and intermediate intellectual performance of adult schizophrenics. *Journal of Consulting Psychology*, 1966, 27, 66-67.
- ARIST, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
- ARIST, S. Schizophrenia. In S. Arast, Ed., *American handbook of psychiatry*, Vol. 1. New York: Basic Books, 1969.
- CHURCH, F. F. A hereditary finding. Sex roles in schizophrenia. *Journal of Abnormal and Social Psychology*, 1964, 69, 392-400.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, 37-46.
- COHEN, L. R. MMPI L-F-K relationships with criteria of behavioral disturbance and social adjustment in a schizophrenic population. *Journal of Consulting Psychology*, 1966, 23, 339.
- HALL, D. O. *The organization of behavior*. New York: Wiley, 1949.
- HALL, D. O. The American revolution. *American Psychologist*, 1966, 15, 735-747.
- LANE, E. & ALDER, G. W. Early childhood intellectual differences between schizophrenic adults and their siblings. *Journal of Abnormal and Social Psychology*, 1966, 69, 187-195.
- LEINSTEIN, I. & NACHREIN, J. *Statistical analysis in educational research*. Cambridge: Riverside Press, 1960.
- LOVE, M., O'CONNOR, J. P. & STERN, J. W. The psychotic reaction profile. *Journal of Clinical Psychology*, 1966, 16, 41-45.
- MILLER, G. A., GALANTER, E. & PIERCE, R. H. *Plans and the structure of behavior*. New York: Holt, 1960.
- PIERCE, R. H. Reinforcement revisited: A critical view. *Nebraska Symposium on Motivation*, 1966, 11, 11-15.
- ROSENTHAL, M. & GREEN, N. Sources of error in visual and auditory perceptions in projective visualization experiments. *Psychological Bulletin*, 1966, 62, 1-13.

(Early publication received October 2, 1968)

EFFECT OF ACTUAL STRESS ON WORD ASSOCIATIONS

SANDOR MINTZ

Colgate University

A Kent Rosanoff Word Association Test was administered to 15 affiliate nurses on the afternoon of the assassination of President Kennedy. About a year later on a less eventful afternoon, a second group of 15 affiliate nurses was given the same test. The first assessment was defined as stressful (the second testing situation was considered as a nonstressor control). The purpose was to note the effects of actual environmental stress on this projective technique. A four-category system of response classification was used based on the Palermo Jenkins college female group. Results indicated a significant decrease in total outcome responses on the stress day but no difference in the frequency of unique responses for the two groups. The stress group also demonstrated less variability in responding.

Since projective tests purport to measure an individual's more permanent, deep seated emotional trends, variations in results due to temporary emotional stress should be limited. Saltz (1961) demonstrated that an induced experimental stress situation adversely affected the responses to the Kent Rosanoff Word Association Test (Kent & Rosanoff, 1915). Goldstein (1961) found no greater frequency of late responses under stress and therefore felt this opposed Sarason's (1961) lower commonality findings under similar conditions.

The assassination of President Kennedy on November 22, 1963 inadvertently provided a situation for studying the effect of actual stress upon this projective technique. On that afternoon a Kent Rosanoff Word Association Test in written form consisting of 15 stimulus words was administered to 15 affiliate nurses at Macy State Hospital, Utica, New York. The group was visibly upset. The identical test was administered to another group of 15 student nurses on January 10, 1964 as a control or nonstressor group. Not only does a comparison of the responses with those obtained on the afternoon of the assassination provide some indication of the effect of actual stress on responses to the word

association test, but a comparison of these results with those of Saltz may also be of heuristic value with respect to the differential effects of actual and experimentally induced stress.

METHOD

Using the Palermo Jenkins (1964) college female norms, a four category system of response classification was developed. If a given response that was given by 10% or more of the norm group it was classified as most common. If a given response that was given by between 7% and 9% of the norm group, it was designated as common. If a given response that was given by between 4% and 6% of the norm group, it was designated as unusual. If a given response that was less than 3% of the norm group, it was designated as unique. For example, one of the stimulus words was "where" in the response word series. This response was given by 10% of the norm group and was therefore considered to be a most common response. In the stressor word sequence, the word "where" was given by 100% which was the response of 100% of the norm group. Therefore, this response was designated common. The stimulus "I" was given by 100% which was the response of 100% of the norm group. The word "I" had the response of 100%. The stimulus "all" was given by 100% which was the response of 100% of the norm group. The word "all" had the response of 100%. The stimulus "and" was given by 100% which was the response of 100% of the norm group. The word "and" had the response of 100%.

RESULTS

The results are displayed in Figure 1. The mean frequency and standard deviation for the stress group were $M = 11.5$, $SD = 1.5$, $t = 1.1$, $p = .27$, and $t = 1.1$, $p = .27$ respectively. The mean frequency and standard deviation for the control group were $M = 11.5$, $SD = 1.5$, $t = 1.1$, $p = .27$, and $t = 1.1$, $p = .27$ respectively. The mean frequency and standard deviation for the stress group were $M = 11.5$, $SD = 1.5$, $t = 1.1$, $p = .27$, and $t = 1.1$, $p = .27$ respectively. The mean frequency and standard deviation for the control group were $M = 11.5$, $SD = 1.5$, $t = 1.1$, $p = .27$, and $t = 1.1$, $p = .27$ respectively.

The author is indebted to William International for assistance in the design and construction of the test. The author is also indebted to the staff of the Macy State Hospital for their assistance in the design and construction of the test. The author is also indebted to the staff of the Macy State Hospital for their assistance in the design and construction of the test.

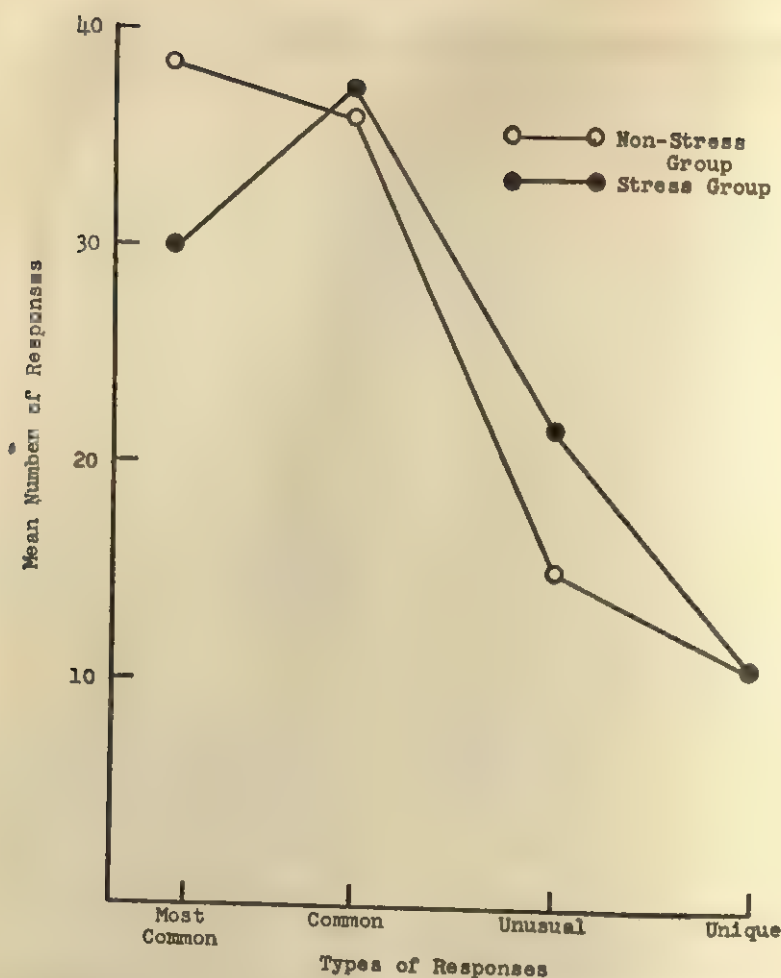


FIG. 1. Comparison of stress and nonstress groups responses on word association test.

dard deviations for the control group were 37.73 ± 11.18 , 36.53 ± 5.14 , 14.73 ± 5.43 , and 11.00 ± 9.74 most common, common, unusual, and unique, respectively. A *t* test demonstrated that the stress and nonstress groups differed significantly on both the most common and the unusual categories of response—the stress group yielding significantly *less* most common responses at an .05 level and significantly *more* unusual responses at an .001 level. However, there was no significant difference in the frequency of common and unique responses between the two groups, the stress group averaging 36 common responses, the nonstress group 37. The mean number of unique responses was identical at 11 for the two groups. Using Cochran's test of Homo-

geneity of Variance a trend toward less variability was noted throughout the four categories in the stress group, the unique category yielding significantly less variability at a .005 level.

DISCUSSION AND CONCLUSION

It appears that certain classes of responses to the word association test are affected by actual stress while others are not. For the practicing clinician it is important to know what these classes are, for such differential responsiveness to situational stress and anxieties must be taken into account when interpreting diagnostic materials.

The unique response, which was demonstrated to be unaffected in frequency by actual

stress, is of primary importance to the diagnostician. That these responses were unaffected by temporary stress would indicate that they may be manifestations of deep seated more permanent emotional trends. The unique response of "hate," for example, to the stimulus word MOTHER is usually more illuminating to the therapist than would be the response "father," which is given by 70% of the norm group. Thus the demonstrated stability of these unique responses indicates their importance within projective testing, and serves as a validating aspect of the word association test.

Saltz (1961) obtained similar results in a study of the effects of experimentally induced stress on the word association test. His findings indicated a highly significant decrement in common responses, yet no tendency toward "novel" responses as a function of his operationally defined stress. This similarity of results would seem to indicate the efficacy of experimentally induced stress in such studies.

Saltz's purpose was to determine whether stress should be considered a stimulus or a drive. He ascertained that if a stimulus, stress should "produce characteristic stress responses which may compete with the common free associations." However, if a drive, stress should "multiply the strength of both common free associations and their competing stress responses. Consequently, the drive characteristics of stress should strengthen the strong common free associations disproportionately more than the weak common associations. Thus, the strong common associations should suffer less of a decrement than the weak common associations." The author's work, although based on actual rather than experi-

mentally induced stress, supports Saltz's finding that the former condition was demonstrated to be the case, indicating the stimulus rather than drive qualities of stress.

In addition, this experiment would seem to offer some resolution to the Goldstein and Sarason discussion. These results, due to the four-category system, seem to offer a resolution for the argument, demonstrating that the lower commonality score may not be due to a greater frequency of rare or unique responses, but to the particular classification system used to categorize responses. Thus, using this system of classification demonstrated that Sarason's lower commonality scores might be attributed to a lowered frequency of most common responses and a greater frequency of unusual responses in the stress situation.

The lowered variability of responding in the experimental group seems to indicate an increase of rigidity precipitated by the stressful situation.

REFERENCES

- GOLDSTEIN, M. J. The relationship between anxiety and oral word association performance. *Journal of Abnormal and Social Psychology*, 1961, 62, 468-471.
- KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *American Journal of Insanity*, 1910, 67, 37-96; 317-390.
- PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college*. Minneapolis: University of Minnesota, 1964.
- SALTZ, E. The effect of induced stress on free associations. *Journal of Abnormal and Social Psychology*, 1961, 62, 161-164.
- SARASON, I. G. A note on anxiety, instructions and word association performance. *Journal of Abnormal and Social Psychology*, 1961, 62, 153-154.

(Received October 27, 1967)

SOCIAL ADJUSTMENT AND FAMILIAL SCHEMA¹

JERRY HIGGINS,² JUDITH C. PETERSON, AND LISE-LOTTE LEHD DOLBY

University of California, Santa Barbara

Two groups of 25 male undergraduates with good and poor social adjustment as determined by a questionnaire freely placed figures of a father, a mother, a son, and a daughter on a field. As predicted, Ss with poor social adjustment placed the son closer to the father, while Ss with good social adjustment placed the son closer to the mother ($p < .01$). Both groups placed the daughter closer to the mother ($p < .01$). The results parallel those obtained from emotionally disturbed versus normal boys.

Kueth (1962a, 1962b, 1964) has demonstrated the existence of common social schemata, that is, fundamental organizing principles in social perception. For example, when male undergraduates are asked to place felt figures of a man, a woman, and a boy on a field, they more frequently place the boy closer to the woman. Weinstein (1965) has recently employed the same technique to investigate the familial schemata of emotionally disturbed boys. Like the majority of Kueth's undergraduates, Weinstein's normal control boys placed the son closer to the mother. The disturbed boys, on the other hand, placed the son closer to the father. Weinstein interpreted these data as reflecting a distant and negative mother-child relationship on the part of the disturbed boys. The present study sought to account for such atypical familial schemata within the "normal" range of functioning. It was hypothesized that whereas well-adjusted normal males exhibit a familial schema in which the son is closer to the mother, less well-adjusted normal males display a familial schema in which the son is closer to the father.

METHOD

One hundred male undergraduates who had lived their entire lives (until entering college) with their biological parents as an intact family unit served as Ss. Adjustment was assessed by a self-administered questionnaire stressing social-sexual interests and activities in high school and college such as friendship patterns, group membership, dating behavior, and so forth (Goodman, 1968). This "social adjust-

ment scale," which was designed for use with college normals, correlates highly with a premorbid adjustment scale (Phillips, 1953) widely used in research in schizophrenia (Garmezy, 1968; Rodnick, 1967). Goodman (1968) administered the scale to 29 schizophrenics whose premorbid adjustment scores were known; a product-moment correlation between the scale scores yielded a coefficient of .75 ($p < .0005$). Possible scores on the scale range from 28 (poor adjustment) to 88 (good adjustment). Extreme quartiles were selected as the poor social adjustment (PSA) and good social adjustment (GSA) groups with mean scale scores of 55.7 and 78.8, respectively. The PSA and GSA groups were similar in age and socioeconomic status.

Kueth's (1962a) Felt Figure Technique was used to determine familial schema. The S was instructed to place four red felt figures (Figure 1) consisting of a man (father), a woman (mother), a boy (son), and a girl (daughter) "in any manner you wish" on a 6 × 3.5 ft. vertically mounted green felt field. The figures were handed to S in a pile, the position of the figures in the pile being randomly varied. Extensive testing in this laboratory has shown that Ss invariably identify the sex and relative age of the figures correctly. The figures were presented



FIG. 1. Felt figures.

¹ This research was supported by a faculty research grant from the University of California to the senior author.

² Request for reprints should be sent to Jerry Higgins, Department of Psychology, University of California, Santa Barbara, California 93106.

TABLE 1
PLACEMENT OF FELT FIGURES

Group	Interfigure distance (in.)						Interventions					
	Total group		Female sibling(s)		No female sibling		Total group		Female sibling(s)		No female sibling	
	Mother-son	Father-son	Mother-son	Father-son	Mother-son	Father-son	Mother-son	Father-son	Mother-son	Father-son	Mother-son	Father-son
PSA												
<i>N</i>	25		17		8		25		17		8	
<i>M</i>	12.22	10.32	14.16	12.07	8.12	6.61	.84	.56	.94	.71	.62	.25
<i>SD</i>	8.76	6.24	9.78	6.73	3.43	2.28	.61	.70	.64	.75	.49	.44
GSA												
<i>N</i>	25		18		7		25		18		7	
<i>M</i>	7.18	9.38	7.00	9.04	7.62	10.25	.36	.80	.33	.78	.43	.86
<i>SD</i>	3.14	3.93	3.53	4.18	1.67	3.04	.56	.69	.57	.62	.50	.83

together to *S*, rather than a pair at a time as in Weinstein's (1965) study, in order to enhance the "family group" percept. The distance between the centers of the heads of the figures was measured to the nearest $\frac{1}{8}$ in. in the *S*'s absence. The *S*'s familial schema was obtained prior to administration of the social adjustment questionnaire.

RESULTS

As may be seen from the left half of Table 1 (total group column), the PSA *S*s placed the son figure closer to the father than to the mother figure, while the GSA *S*s placed the son figure closer to the mother than to the father figure—Groups \times Parents interaction, $F = 7.29$, $df = 1/48$, $p < .01$. In terms of relative frequency, 76% of the GSA group placed the son figure closer to the mother than to the father figure, whereas only 40% of the PSA group displayed this schema ($\chi^2 = 5.25$, $df = 1$, $p < .025$). Had the "poor adjustment" end of the assessment scale been better represented, it seems probable that the contrast between the groups would have been even sharper.

It should be noted that the above differences in familial schema between the PSA and GSA groups were not merely child-parent differences, but were sex specific for the child figure of the same sex as *S*s, that is, the son figure. The combined groups placed the daughter figure closer to the mother than to the father figure—parents effect, $F = 16.39$, $df = 1/48$, $p < .01$, with the Groups \times Parents interaction not approaching significance— $F < 1$.

"Closeness" between son and a given parent

figure may be viewed not only in terms of interfigure distance, but also in terms of the number of other figures (i.e., other parent and daughter) placed between the son and that parent figure. The right half of Table 1 (total group column) reveals that the PSA group placed a greater number of such intervening figures between the son and mother figures than between the son and father figures, whereas the GSA group more frequently placed interventions between the son and father figures than between the son and mother figures—Groups \times Parents interaction $F = 7.20$, $df = 1/48$, $p < .01$.⁸ This was true whether the intervening figure was that of the other parent—Groups \times Parents interaction $F = 4.50$, $df = 1/48$, $p < .05$, or the daughter—Groups \times Parents interaction $F = 4.04$, $df = 1/48$, $p = .05$. The same trend was reflected in the proportions of PSA and GSA *S*s placing more interventions between son and mother figures, no interventions between son and either parent figure, and more interventions between son and father figures ($\chi^2 = 9.16$, $df = 2$, $p < .02$). Again, these differences in schemata proved sex specific for the son figure, since the combined groups placed fewer interventions between the daughter and mother figures than between the daughter and father figures—parents effect $F = 23.44$, $df = 1/48$, $p < .01$, the Groups \times Parents interaction falling far short of significance— $F < 1$.

⁸ Keuthe (1962a) also reported that the schema in which the man was placed between the woman and boy was relatively rare.

Since a daughter figure was used with all Ss, subanalyses were conducted with respect to whether or not the families of the Ss actually contained a female sibling (Table 1). While all such analyses yielded results which lay in the predicted direction, only that which treated interfigure distance for Ss with female siblings reached the .05 level of significance, probably owing to the reduction in N .

DISCUSSION

The results indicate that normal males who differ in the adequacy of their social adjustment also differ in their familial schema. The familial schema of well-adjusted males includes greater proximity of the male child to the mother, with fewer family members permitted to come between mother and son. Relatively poorly adjusted males, however, possess a familial schema of greater distance and more interventions between mother and son. The relative position of the male child in the poorly adjusted male's schema represents a sex-specific deviation, since both poorly and well-adjusted males maintain close proximity between female child and mother. Inasmuch as "The schema that places a child with a woman . . . is probably one of the first specific social schemata developed by most children [Kuethe, 1964, p. 253]," and insofar as the physical distance placed between human figures may be equated with emotional distance (Weinstein, 1965), the results may reflect a negatively toned maternal relationship in the early lives of males whose social-sexual adjustment is poor—at least as perceived by the child.

An alternative explanation of the results might argue that the obtained data reflect identification with the father on the part of the PSA Ss, and with the mother by the GSA Ss. However, when asked which parent they "feel most like" (perceived similarity), there was a nonsignificant tendency for the PSA group (32%) to report "mother" more frequently than the GSA group (16%). Further, from the very nature of the adjustment questionnaire, in which items deal with social and sexual activity and assertiveness, it is apparent that the sex typing of the GSA group is more masculine than that of the PSA group. Although these counterarguments do not defin-

itively refute an identification hypothesis, they do make unlikely any simple explanation of the PSA and GSA schemata on the basis of identification.

A final problem relates to possible influences of a social desirability set (Crowne & Marlowe, 1964). However, when the Marlowe-Crowne Social-Desirability Scale was correlated with the social adjustment scale in this laboratory, the resulting coefficient was small ($r = -.31$, $df = 22$, $p > .10$). Further, evidence of the distorting effects of social schemata on judgment indicates that the schemata are not "merely the result of subjects attempting to respond the way they believed was 'correct' or the way 'most people' would respond [Kuethe, 1962b, p. 74]." Thus, explanation of the results in terms of social desirability would seem somewhat improbable.

Using the same social adjustment scale as the present study in a two-choice decision situation, Goodman (1968) found that PSA Ss responded more slowly when influenced by a maternal than by a paternal figure, while GSA Ss responded with more nearly equal latencies to maternal and paternal figures.⁴ Since long latencies have been reported to be indexes of conflict (Rotter, 1951), Goodman's results appear consonant with those of the present study. The present results also parallel those of Weinstein (1965) obtained from emotionally disturbed boys. This parallelism is of importance not only in the fact that similar differences in familial schemata hold at different age levels, but also that their appearance may be predicted within the normal population as well as between normal and pathological groups. Familial schemata are apparently sensitive to relatively subtle variations in adjustment and, by inference, early parent-child relationships.

⁴ The caption under Goodman's Figure 3 (p. 509) is incorrect and should be replaced by the caption under Figure 4 (p. 510).

REFERENCES

- CROWNE, D. P., & MARLOWE, D. *The approval motive: Studies in evaluative dependence*. New York: Wiley, 1964.
- GARMEZY, N. Process and reactive schizophrenia: Some conceptions and issues. In M. M. Katz, J. O. Cole, & W. E. Barton (Eds.), *The role and*

- methodology of classification in psychiatry and psychopathology*. (No. 1584) Washington: Public Health Service Publication, 1968.
- GOODMAN, I. Z. Influence of parental figures on schizophrenic patients. *Journal of Abnormal Psychology*, 1968, 73, 503-512.
- KUETHE, J. L. Social schemas. *Journal of Abnormal and Social Psychology*, 1962, 64, 31-38. (a)
- KUETHE, J. L. Social schemas and the reconstruction of social object displays from memory. *Journal of Abnormal and Social Psychology*, 1962, 65, 71-74. (b)
- KUETHE, J. L. Pervasive influence of social schemata. *Journal of Abnormal and Social Psychology*, 1964, 68, 248-254.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- RODNICK, E. H. Cognitive and perceptual response set in schizophrenics. In R. Jessor & S. Feshbach (Eds.), *Cognition, personality, & clinical psychology*. San Francisco: Jossey-Bass, 1967.
- ROTTER, J. B. Word association and sentence completion methods. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques and other devices for understanding the dynamics of human behavior*. New York: Prentice-Hall, 1951.
- WEINSTEIN, L. Social schemata of emotionally disturbed boys. *Journal of Abnormal Psychology*, 1965, 70, 457-461.

(Received November 13, 1967)

APPROACH AND AVOIDANCE RESPONSES OF THE HYSTERICAL PERSONALITY TO SEXUAL STIMULI¹

MARION O'NEILL² AND BERNHARD KEMPLER³

Emory University

Thirty-two female hysterical personalities and 32 nonhysterical personalities performed a paired-associates learning task and a visual recognition task. Both tasks included neutral and sexual material. Half the Ss in each group performed under sexually "seductive" conditions, the other half under sexually neutral conditions. Results showed that (a) hysterics learned sexual material quickly under neutral conditions but relatively slowly under sexual conditions; (b) hysterics had a relatively low visual recognition threshold for sexual words under neutral conditions and a relatively high threshold under sexual conditions; and (c) these threshold differences were maximized when sexual stimuli referred to the self. It was concluded that the characteristic behavior of the hysteric was consistent with an approach-avoidance conflict involving sexual motives and sexual behavior.

The earliest references to the "hysterical personality" (e.g., Jaspers and Janet as quoted by Muncie, 1939) were descriptions of the personality characteristics of individuals with conversion symptoms. More recently several writers have advanced the concept of hysteria as a personality dimension, even for individuals who show no evidence of conversion symptoms (Chodoff & Lyons, 1958; Fidern, 1940; Fitzgerald, 1948; Schilder, 1939; Thornton, 1948). With the exception of psychoanalytically oriented case studies and a few studies testing the hypothesis of the greater suggestibility of conversion hysterics, the literature on the hysterical personality has not been extensive. The absence of an objective definition, as well as the difficulties in testing the psychoanalytic hypotheses concerning fixation at different psychosexual stages, has contributed to the lack of empirical work in this area.

The present study is concerned with the frequently observed combination in the female

hysteric⁴ of sexually "provocative" behavior with the denial of sexuality. Chodoff and Lyons (1958) in their review of the literature on the hysterical personality found frequent mention, on the one hand, of "lasciviousness, sexualization of all non-sexual relations, coquetry, and provocativeness," and, on the other hand, "sexual frigidity, intense fear of sexuality and failure of the sex impulse to develop toward its goal." According to Lewis (1956), hysterics are both "coquettish and frigid." Laing (1962), writing from an existentialist viewpoint, stated that "sexuality is very here and now and so the hysteric flirts with it as she flirts with life but without ever being able to face it." Starr (1953) mentioned as one of the hysteric's characteristics the "persistence in sexualizing all non-sexual relations," and Marmour (1953), arguing for an oral-fixation interpretation, stated that the "hysteric is distressed when her sexuality brings on advances, as she is really asking to be loved as a child and not as a woman."

These interpretations, in spite of their varied theoretical perspectives, appear to be descriptive of a sexual approach-avoidance conflict. When relatively far from the sexual goal, the behavior of the hysteric appears to be sexually motivated and to have socially

⁴ The status of the hysterical personality in the male is, as yet, ambiguous and requires a more extended discussion than is called for in this paper.

¹ This investigation was supported in part by a Public Health Service Fellowship No. 5-FI-MH-24, 454-02 (MTLH) from the National Institute of Mental Health and is based on an unpublished doctoral dissertation carried out by Marion O'Neill under the direction of Bernhard Kempler.

² Now at the Georgia Mental Health Institute.

³ Requests for reprints should be directed to Bernhard Kempler, who is now at the Department of Psychology, Georgia State College, Atlanta, Georgia 30303.

defined sexual significance. When close to the sexual goal, her behavior is characterized by avoidance of sexual stimuli and denial of sexual feelings and motives. It can be further assumed that this approach-avoidance conflict is evident not only in the interpersonal behavior of the hysteric but on the level of perceptual and cognitive processes as well. Thus, under relatively neutral (i.e., non-sexual) conditions, the hysteric should be particularly sensitive to and observant of stimuli which have sexual implications. However, under sexually "provocative" conditions she should be selectively unaware of sexual stimuli, particularly if a relationship of these stimuli to her is implied.

The present study tested these hypotheses. Two groups of female Ss, hysterics and non-hysterics, were required to perform a paired-associates learning task and a visual recognition threshold task. Half of the Ss in each group worked under sexually neutral conditions, the other half under sexually "provocative" conditions. Both tasks involved neutral and sexual verbal material. In addition, the manner of presentation of the stimuli in the recognition threshold task encouraged either a set toward self-reference or a set toward other-reference. The specific predictions were (a) under sexually neutral conditions hysterics, compared with nonhysterics, would make relatively fewer errors in paired-associate learning of sexual words and would have a lower recognition threshold for sexual words, and (b) under sexually "provocative" conditions hysterics, compared with nonhysterics, would make relatively more errors in paired-associates learning of sexual words and would have a higher threshold for sexual words, particularly when self-reference of these words is implied.

METHOD

Subjects

Female undergraduate students ($N = 150$), enrolled in a physical education course, were administered the Edwards Personal Preference Scale (EPPS) and the Guilford-Zimmerman Temperament Survey (G-Z). Previous work by O'Neill⁶ has shown that,

⁶ M. O'Neill, unpublished master's thesis, "Validity of Hysterical Personality as a Clinical Category," 1965.

TABLE 1
FOUR CATEGORIES OF STEM AND RESPONSE
WORDS

Category			
H-H	H-L	L-L	L-H
Kiss C_ck	Caress B_st	Book R_te	Write L_ve
Penis B_lls	Petting W_rk	Walk N_te	Swim D_ck
Screw C_me	Embrace P_ge	Study M_re	Running S_ck
Rape F_ck	Breast J_st	Read B_ck	Desk C_nt

Note.—H-H = sexual stem-sexual response; H-L = sexual stem-nonsexual response; L-L = nonsexual stem-nonsexual response; L-H = nonsexual stem-sexual response.

using a discriminant-function analysis, the combined discriminant weights of six scales (Abasement, Aggression, and Change from the EPPS; General Activity, Personal Relations, and Masculinity-Femininity from the G-Z) yielded the best discrimination between hysterics and nonhysterics as diagnosed by psychotherapists. In a follow-up validity study O'Neill⁶ showed that female nonpatient groups, scoring high on this hysterical personality measure were judged by clinicians in interviews to have hysterical personality traits, while Ss scoring low were judged to have significantly fewer hysterical personality traits.

Next, 56 Ss were selected from *S* pool, 28 defined as hysterics and 28 as nonhysterics, based on their extreme scores on O'Neill's hysterical personality measure (HP). The two groups were also equated for verbal learning ability on a 5-word nonsense-syllable task. One-half of the hysterics and one-half of the nonhysterics were assigned to the experimental (sexually "seductive") group and the remaining Ss to the control (sexually neutral) group.

Procedure

All Ss were tested individually by a male examiner.⁷ Each *S* was administered a paired-associates learning task (PAL) and a visual recognition threshold (VRT) task. One-half of the Ss received the PAL first while the other half received the VRT first.

PAL. Stimuli consisted of 16 stem words paired with 16 incomplete response words. Eight words were selected for their sexual meaning; the other 8 were selected for activity connotation which was not sexual. The incomplete response words which had previously been rated by college females for their degree of sexual connotation were words of four and five letters with the second letter omitted. Four sexual words were randomly paired with 4 sexual response items; the other 4 sexual words were randomly paired with nonsexual items. The same procedure was followed for the nonsexual stems and

⁶ M. O'Neill, unpublished study entitled, "Validation Study for Hysterical Personality Scales," 1966.

⁷ The authors are indebted to William Ewald and Brian Jordan for their help in this research.

THEY READ

I KISS

FIG. 1. Sample of pronoun and verb lettering size.

response items. Table 1 shows the four categories of stem and response items.

Stimuli were projected onto a 4 × 5 ft. standard movie screen by a Kodak Carousel No. 800 projector equipped with an automatic timing device. The projector was mounted on a table located behind and to the left of *S*; the screen was located 3 ft. from *S*. Each stimulus was presented for 5 sec. The *study-test method* of presentation was used, each *S* receiving a training trial followed by a test trial. The training trial showed each stem word alone followed by presentation of the stem word paired with the correct response item. The test trial presented the stem word only. It was *S*'s task to spell out loud the letters of the incomplete response word correctly paired with the stem word. Any error on the test trial necessitated a complete training trial. A 2-sec. interval separated the training from the test trial. Criterion of mastery was 100% correct recall of the response list (i.e., one errorless trial).

VRT. The stimuli were 12 two-word phrases, consisting of six verbs, each one combined with the pronoun, *I* (self-reference), or with the pronoun, *They* (other-reference). Three verbs were sexual in content (rape, kiss, screw) and three were neutral (read, swim, study). All verbs appeared in the PAL task. Procedure and method of presentation were similar to those reported by Kempler and Wiener (1964). The pronouns were printed with a thick pen point (Leroy No. 4) and the verbs with a thin pen point (Leroy No. 0). When presented in the tachistoscope the pronouns were more distinct and easier to recognize than the verbs. Figure 1 shows two sample phrases.

A two-channel, Dodge-type, electronically triggered tachistoscope, No. 800 F was used. Recognition threshold was established by the method of ascending limits. The first exposure for each *S* began at the lowest Vernier-scale setting and was increased by half-unit steps thereafter until *S* correctly identified the stimuli twice. The 12 phrases were presented in random order to each *S*.

Experimental and Control Conditions

All *Ss* were tested in the evening. They were met by a male examiner in a waiting room and taken by him to an adjoining experimental room. This room was in semidarkness when *S* entered and

remained so throughout the experiment. It was divided into two sections by a partition behind which sat a female assistant who recorded *S*'s responses.

Two male examiners were used, each one testing half of the *Ss* in the hysteric group and half in the nonhysteric group. The *Es* had no knowledge of the diagnostic classification of the *Ss*. Both men were tall, in their 20s, and generally considered to be handsome. The experimental and control conditions were defined by *E*'s behavior toward *S*. In the experimental (sexually "seductive") condition the examiner (*a*) made personal and flattering comments to *S* (e.g., "I like your dress"), (*b*) made at least one physical contact with *S* prior to the beginning of the experiment (e.g., taking her arm while leading her to the experimental room), and (*c*) upon entering the experimental room locked the door in order to "insure privacy." In addition both examiners, by their tone of voice and other subtle methods, behaved "seductively" but not in a crude or offensive manner. In the control (sexually neutral) condition the examiner was friendly but impersonal and business-like in his comments to *S*, made no physical contact with her, and did not lock the door of the experimental room.

RESULTS

Paired-Associates Learning Task

The measures obtained for the PAL were errors to criterion. A three-factor with repeated measures analysis of variance (Winer, 1962) was computed on these measures, the three factors being personality type, experimental condition, and category of stem-response item combination. Results of the analysis are shown in Table 2.

The main effect of *S* groups was not signifi-

TABLE 2
PAIRED-ASSOCIATE LEARNING ANALYSIS OF VARIANCE
FOR NUMBER OF ERRORS ON FOUR
CATEGORIES OF WORDS

Source	df	MS	F
Between	55		
Groups	1	11.161	.557
Conditions	1	412.572	20.609**
Groups × Conditions	1	261.446	13.059**
Error	52	20.019	
Within	168		
Categories	3	91.583	19.691**
Groups by Category	3	4.744	1.019
Conditions × Category	3	27.654	5.945*
Groups × Conditions × Category	3	18.458	3.968*
Error	156	4.65	

* $p < .01$.

** $p < .001$.

cant, indicating that, other things being equal, hysterics and nonhysterics do not differ in paired-associates learning ability. The main effects of experimental conditions ($p < .001$) and of word categories ($p < .01$) were significant, as were the interactions of groups with conditions ($p < .001$) and groups with verbal categories ($p < .01$). These effects are most clearly revealed by examining the significant ($p < .01$) triple interaction shown in Figure 2. It is clear that the "seductive" behavior of the examiner had little effect on the nonhysterics, but had a marked effect on the hysterics. Furthermore, this effect is related to the degree of sexual meaning contained in the verbal categories.

Visual Recognition Threshold

VRT was defined as the number of constant increments of exposure time required for *S* to correctly report twice each of the 12 stimulus phrases. The threshold data were transformed into difference scores (D scores) by subtracting each *S*'s *I* sexual threshold scores from her *They* sexual threshold scores, and her *I* neutral from her *They* neutral threshold scores. Table 3 shows the results of a three-factor analysis of variance with repeated measures (Winer, 1962). The only significant effects were the main effects of word categories ($p < .05$), and the triple interaction ($p < .05$). Comparison of means (Scheffé, 1959) showed that the mean D score for sexual words under experimental conditions

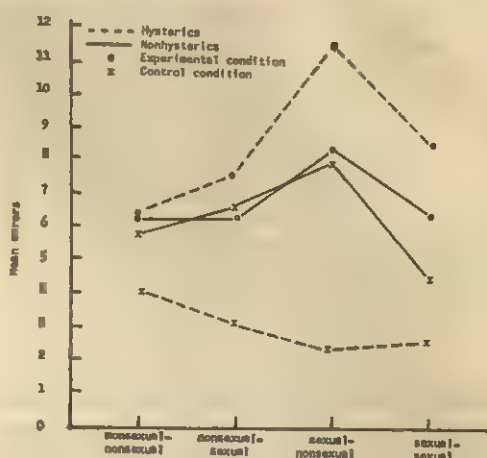


FIG. 2. Paired-associate learning task. (Errors to criterion under experimental—sexually seductive—and control—sexually neutral—conditions for four categories of word stem-response combinations.)

was significantly higher ($p < .01$) for the hysterics than any other group (see Table 4).

DISCUSSION

The results of this study are in full agreement with the hypotheses, and offer support for the interpretation of at least some aspects of the behavior of the hysterical personality as a sexual approach-avoidance conflict. Under nonsexual conditions the hysteric is sensitized to stimuli with possible sexual connotations. Conversely, under sexually seductive conditions she avoids or selectively inattends to stimuli with sexual connotations.

The dimension of "distance from sexual goal" is, no doubt, a complex one. The procedures of this study permitted the differentiation of two meanings of this concept: (a) the objective distance, that is, the culturally defined sexual implications of the physical setting and of interpersonal behaviors, and (b) the subjective distance, that is, the awareness and acceptance of one's own sexual motives. The latter meaning of distance was represented in the VRT by the self-reference and other-reference pronouns, *I* and *They*, respectively. The data indicate that these two aspects of distance may interact; that is, the hysteric is aware and accepting of sexual self-reference when the environmental conditions are sexually irrelevant or neutral. How-

TABLE 3

VISUAL-RECOGNITION TASK ANALYSIS OF VARIANCE FOR SEXUAL AND NONSEXUAL WORDS

Source	df	MS	F
Between	55		
Groups	1	650.893	1.21
Conditions	1	282.893	
Groups \times Conditions	1	641.286	1.19
Error	52	538.09	
Within	56		
Verb category	1	1,586.286	4.91*
Groups \times Verbs	1	408.893	1.35
Conditions \times Verbs	1	282.893	
Groups \times Conditions \times Verbs	1	1,856.571	6.14*
Error	52	302.56	

* $p < .05$.

TABLE 4

VISUAL RECOGNITION THRESHOLD MEAN TRIAL
D-SCORE FOR SEXUAL AND NONSEXUAL
WORDS

Group	Threshold value	
	Sexual words	Nonsexual words
Hysterical Ss	10.92	-11.50
Hysterical controls	-8.4	- 8.15
Nonhysterical Ss	-1.00	1.50
Nonhysterical controls	4.57	-10.72

ever, when the environmental conditions are rich in sexual implications she denies and avoids sexual self-reference, while at the same time becoming highly aware of sexual motives and behaviors in others.

Previous studies in the personality-perception area with emotion or affect arousing words have been fraught with conceptual and methodological difficulties (e.g., Goldiamond & Hawkins, 1958; Goldstein, 1962; Howes & Solomon, 1951; Solomon & Postman, 1952). While the results of this study demonstrate that learning and recognition threshold behavior of hysterics are affected by sexually "seductive" conditions, they do not offer an unequivocal basis for the description of the processes responsible for this effect. For example, the interpretation of response suppression (i.e., "seeing" but not "saying") has sometimes been offered for studies using taboo or "dirty" words (Howes & Solomon, 1950). The affect-arousing words in this study, while clearly sexual in meaning, were not of the "dirty" variety; nevertheless, the response-suppression interpretation could still be relevant here, particularly in view of the female Ss and the male Es.

The position taken here is that when the learning and recognition of classes of words are predictably affected in different Ss by verbal (e.g., *I* and *They*) as well as interpersonal context, such data are relevant for personality whether one invokes special and unusual processes (e.g., subception, perceptual defense) or makes use of more mundane explanations, such as response suppression due to embarrassment or prudishness. One need only assume that with specific cues differences in personality will show up as different response characteristics in the pres-

ence of such cues (Kempler & Wiener, 1963, 1964), and furthermore, that response characteristics are affected by a variety of conditions—physical, social, and interpersonal.

Such an orientation would not be expected to yield much information about the intrapsychic characteristics of different personalities, but it would be productive of statements concerning reactions of certain personalities to classes of cues under particular conditions. The hysterical personality, then, is one who, under sexually neutral conditions, sensitizes to a large variety of sexual cues, and who, under sexually "provocative" conditions, selectively inattends to the same sexual cues, particularly when self-reference is implied.

Other characteristics of the hysterical personality, leading to a more comprehensive description, can be investigated by other combinations of cues in different (physical, interpersonal, etc.) settings.

REFERENCES

- CHODOFF, P., & LYONS, H. Hysteria: The hysterical personality and hysterical conversion. *American Journal of Psychiatry*, 1958, 114, 734-740.
- FEDERN, B. The determination of hysteria versus obsessional neurosis. *Psychoanalytic Review*, 1940, 27, 265-276.
- FITZGERALD, O. W. Love deprivation and the hysterical personality. *Journal of Mental Science*, 1948, 94, 701-717.
- GOLDIAMDOND, I., & HAWKINS, W. F. Vexiersuch: The log-relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology*, 1958, 56, 456-463.
- GOLDSTEIN, M. I. A test of the response-probability theory of perceptual defense. *Journal of Experimental Psychology*, 1962, 63, 23-28.
- HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "emotionality and perceptual defense." *Psychological Review*, 1950, 57, 229-234.
- HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, 1951, 41, 401-410.
- KEMPLER, B., & WIENER, M. Personality and perception in the recognition threshold paradigm. *Psychological Review*, 1963, 70, 349-356.
- KEMPLER, B., & WIENER, M. Personality-perception: Characteristic responses to available part-cues. *Journal of Personality*, 1964, 32, 57-74.
- LAING, R. D. *The self and others: Studies in existential analysis and phenomenology*. Chicago: Quadrangle Books, 1962.
- LEWIS, A. Hysteria. In D. Hunter (Ed.), *Price's textbook on the practice of medicine* (9th ed.). London: Oxford Universities Press, 1956.

- MARMOUR, J. Orality in the hysterical personality. *Journal of American Psychoanalysis*, 1953, 1, 656-671.
- MUNCIE, W. *Psychology and psychiatry*. St. Louis: Mosby, 1939.
- SCHEFFÉ, H. The analysis of variance. In W. C. Guenther (Ed.), *Concepts of statistical inference*. New York: McGraw-Hill, 1965.
- SCHILDER, F. The concept of hysteria. *American Journal of Psychiatry*, 1939, 95, 1389-1413.
- SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, 1952, 43, 195-201.
- STARR, P. Some observations on the diagnostic aspects of childhood hysteria. *Nervous Child*, 1953, 10, 214-230.
- THORNTON, N. What correlation is there between hysteria and cyclothymia or schizothymia? *Journal of Clinical Psychopathology*, 1948, 2, 537-560.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received November 17, 1967)

SIZE ESTIMATION IN SCHIZOPHRENIA AS A FUNCTION OF CENSURE, DIAGNOSIS, PREMORBID ADJUSTMENT, AND CHRONICITY

PETER A. MAGARO¹

Northern Illinois University

Schizophrenics divided in terms of chronicity, diagnosis, and premorbid adjustment estimated the size of a line embedded in different affective and neutral content. The main result was that premorbid adjustment and chronicity interacted in size-estimation performance with the good, premorbid-adjustment group exhibiting a smaller size estimation in the chronic condition than in the acute, while the poor, premorbid-adjustment group exhibited the opposite effect. Although a Premorbid Adjustment \times Chronicity interaction was predicted in previous work, the specific over- or underestimation tendency found for each schizophrenic subgroup in the present study differed from those previously predicted. The differences in the present results and predictions derived from a review of the literature were discussed in terms of possible difficulties inherent in making inferences from studies which did not directly control the relevant subject dimensions. The results indicated little support for the social-censure hypothesis and suggested the greater effect of organismic variables relative to stimulus variables in predicting schizophrenic task performance.

Rodnick and Garmezy (1957) have proposed a social-motivational model which has emphasized the motivational qualities of specific stimulus conditions in the determination of the schizophrenic-performance deficit. It is proposed that the schizophrenic has undergone early censorious child-rearing experiences which have sensitized him to cues depicting censure. The deficit is due to censorious stimuli eliciting anxiety to which there has been learned a characteristic withdrawal response which interferes with task performance (Dunham, 1959; Magaro, 1967). Other investigators have not found a relationship between censure and performance (Klein, Cicchetti, & Spohn, 1967), and others have considered the deficit to be related to any affective stimulus (Buss & Lang, 1965).

Silverman (1964) has emphasized the relationship between the schizophrenic performance deficit and particular subject dimensions. The deficit is considered to be related to the attention-response disposition characteristic of each schizophrenic subgroup rather than to the effects of particular stimuli. The attention-

response disposition of interest to the present paper is that of scanning which is considered to be more characteristic of certain schizophrenic subgroups than others. The dominance of the scanning disposition for a particular subgroup would determine the over- or underestimation tendency in a size-estimation task. Stimulus content such as censure is considered to be anxiety inducing as in the above social-motivational model, but it is considered to exaggerate the response disposition of a schizophrenic subgroup rather than elicit a competing response of withdrawal. Silverman's hypothesis regarding the scanning disposition of schizophrenic subgroups dichotomized on the basis of premorbid adjustment and chronicity has been derived from the results of separate studies using different combinations of the two subject dimensions. These studies differed in the tasks used, the means of categorizing the two dimensions, and the information available regarding the presence of the chronic and premorbid subject dimensions as well as other possible confounding subject dimensions such as that of diagnosis. This lack of specification of what subject variables are operating in the experimental situation makes inferences con-

¹Requests for reprints should be directed to Peter A. Magaro, Department of Psychology, University of Maine, Orono, Maine 04473.

cerning the effects of premorbid adjustment and chronicity fairly speculative. The present study examined the effects of chronicity and premorbid adjustment on a size-estimation task using both paranoid and nonparanoid groups. From the inferences of Silverman, an interaction between chronicity and premorbid adjustment would be predicted. Acute poors and chronic goods should overestimate, and chronic poors and acute goods should underestimate. Paranoids in each group should exhibit greater underestimations than nonparanoids. It would also be predicted that the under- or overestimation tendency of each group will be exaggerated under censure conditions.

The general purpose of the present study was to contrast aspects of the social-censure model of Rodnick and Garmezy (1957) with what is considered the organismic model of Silverman (1964). Although both positions predict that censorious cues will have drive properties, the social-censure model considers differences in drive to be mainly responsible for schizophrenic subgroup differences. The presentation of specific censorious cues increases drive and consequently the response tendency of withdrawal. Therefore, differences between schizophrenic subgroups would be predicted only under specified stimulus conditions.

The organismic model views the schizophrenic-performance deficit as being mainly related to the habit system. Differences in subgroup performance are due to differences in the habit strength of the response disposition of scanning. In a size-estimation task, the relative performance of two schizophrenic subgroups depends upon the dominance of scanning in the response hierarchy. More specifically, the organismic model would predict differences between groups rather than between stimuli.

An attempt was also made to assess the deficit in size estimation in terms of a response to an immediate presentation of the stimulus versus memory of the stimulus. Yates (1966) suggested that the schizophrenic deficit was mainly a short-term memory function rather than a distortion in the reception of the stimulus. Previous studies utilized a short-term memory procedure (Harris, 1957;

Zahn, 1959), and, therefore, it seemed advisable to compare performance on a direct, size-estimation procedure to one involving size estimation in a memory procedure.

METHOD

Subjects

The sample consisted of 72 hospitalized, schizophrenic males, mostly veterans, ranging in age 20-60 yr. An attempt was made to equate for race, social class of father (Hollingshead, 1960), education, urban versus rural, age, and drug intake. All drugs were phenothiazine derivatives with the vast majority of patients being given Thorazine. All Ss lived with their mothers at least until the age of 16 and had been diagnosed as schizophrenic without complications of organicity, alcoholism, or physical handicaps. The patients meeting the above criteria were divided into equal groups of chronic (5 yr. or more of total hospitalization) and acute (3 yr. or less). These groups were subdivided into good (a score of 14 or less) and poor (15 or more) levels of premorbid adjustment as derived from the Phillips Scale of Premorbid Adjustment (Phillips, 1953). Each subgroup of 18 Ss was further divided into 6 paranoids and 12 nonparanoids according to both the hospital's admitting and current diagnosis. Twelve control Ss were composed of aides and middle-class staff members.

Materials

The relevant cue was a 228-mm. vertical black line embedded in different contextual stimulus scenes. Two \times two transparencies of the stimulus scenes were projected from a Kodak Carousel slide projector to a glass-beaded screen. The size of the drawing paper was 20 \times 25 in.

The practice stimulus consisted of a right triangle with the vertical line emphasized by being black and heavy in contrast to the base and its hypotenuse. The neutral-content stimuli were a single, neutral, black line and a single, vertical, black line between two trees. A stimulus denoting positive affect was a single, vertical, black line between a mother and son in which the mother is giving her son some milk from a pitcher. A censure stimulus was a single, vertical, black line between a son and mother in which the mother is scolding her son. The same scenes without the line were those used by Harris (1957).

Procedure

The procedure was as follows. Each S was seated at a small table with paper and pencil before him and the screen 10 ft. to the front. The S was allowed at least 5 min. to acclimate to the room and the equipment. The practice stimulus was presented with the following directions: "See this triangle? Now look at this line; it is darker and heavier than the others. Draw it on the paper in front of you the same length as you see it here. Let me

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR CONTROLLED AND VARIED SUBJECT VARIABLES

Variable	Subgroup				
	AG	AP	CG	CP	N
Age (in years)					
<i>M</i>	36.2	35.6	48.3	43.8	39.4
<i>SD</i>	17.37	5.87	11.06	10.27	14.04
% urban	100	92	83	83	83
% Classes IV & V (lower class)	100	83	83	67	75
Education (in years)					
<i>M</i>	11.60	9.67	9.91	10.50	12.8
<i>SD</i>	1.88	4.18	2.60	2.14	3.73
Phillips score					
<i>M</i>	11.0	22.6	8.9	20.8	
<i>SD</i>	2.08	3.51	3.75	4.80	
Total hospitalization (in years)					
<i>M</i>	1.56	1.45	11.22	11.25	
<i>SD</i>	1.34	.81	8.30	6.06	
Medication (mgs. per day) (phenothiazine derivatives)					
<i>M</i>	223.92	216.58	208.75	311.08	
<i>SD</i>	227.07	152.42	247.00	322.62	

Note.—AG = acute goods; AP = acute poors; CG = chronic goods; CP = chronic poors; and N = normals.

know when you are finished. Now I am going to hold it up again, but this time I will show it to you for 5 sec., then I will turn it over. After I turn it over, draw a line from memory." If *S* indicated that he understood the task by his performance, *E* then went on to the first experimental trial which was a single black line. If *S* seemed to be having difficulty, the practice trial was repeated.

The directions for the experimental scenes were as follows: "Now I am going to show you a picture of a black line on this screen in front of you. When the picture goes on, draw a line on this paper the same size as the one on the screen." Then, without comment, *E* provided a fresh sheet of paper and gave instructions from memory. After the presentation of the line stimulus, the other content stimuli were presented with the following directions: "Now I am going to show you a picture of a black line between two trees, draw the line the same size as you see it." (Then instructions were given for reproduction from memory.) "Now I am going to show you a picture of a black line between a mother who is scolding her son, she is really bawling him out. Draw the black line the same size as you see it." (Then *Ss* were instructed to reproduce from memory.) "Now I am going to show you a picture of a black line between a mother who is feeding her son. She is giving him a glass of milk." (Then instructions were given for reproduction from memory.)

Within each *S* group, the order of stimulus pre-

sentation was randomized except for the line stimulus which was always first. GSR and base rate were also recorded at this time. These results will be reported in a separate paper.

RESULTS

Table 1 presents the means and standard deviations for controlled and varied subject variables. The acute-chronic and poor-good grouping separated *Ss* as intended. The only other variable which showed a significant difference was medication ($\chi^2 = 28.47$, $df = 3$, $p < .01$). Chronic poors exhibited the greatest difference by being at a high level of medication. The correlation between amount of medication and size estimation was not significant for the total group of patients nor for any single group. When those *Ss* who were not receiving any medication at all were eliminated from the analysis, there was a significant negative correlation between size estimation and amount of medication for the nonparanoid *Ss* ($r = -.44$, $df = 32$, $p < .01$). This effect was mainly due to the nonparanoids in the chronic good ($r = -.86$, $df = 5$, $p < .01$) and acute good ($r = .91$, $df = 6$,

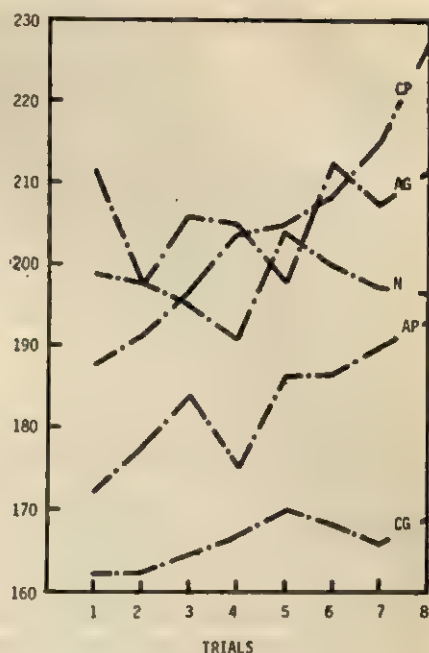
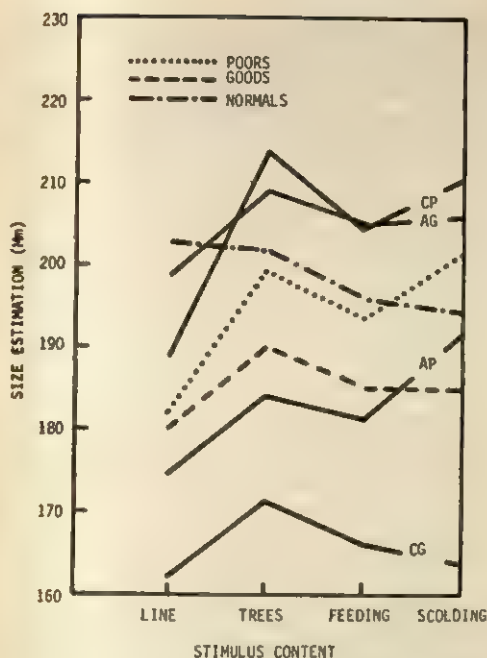


FIG. 1 PLOTS OF THE MEAN SIZE ESTIMATION FOR EACH SCHIZOPHRENIC SUBGROUP AND NORMALS FOR CONTENT AND FOR TRIALS (SEQUENCE)

$p < .01$) subgroups; for these Ss the smaller the amount of medication, the larger the size estimation. The total amount of medication taken by paranoids and nonparanoids did not differ.

The first analysis was a repeated measure four-way (poor-good, chronic-acute, stimulus content, and memory-reproduction) analysis of variance (Winer, 1962). The significant effects were stimulus content ($F = 6.59$, $df = 3/204$, $p < .01$) and the Poor-Good \times Chronic-Acute interaction ($F = 4.52$, $df = 1/68$, $p < .05$). The insignificant memory-reproduction effect indicated that presentation and memory procedures do not have differential effects and only seem to act as additional trials for each stimulus. In examining the content effect, it seemed that the main difference between stimuli was due to a small size estimation on the first-presented nonrandomized stimulus (black line) and a larger but similar size estimation on the other randomized stimuli, (trees, feeding, scolding). This suggested that the content effect was due to order of presentation (sequence). Consequently, a three-way (poor-good, chronic-acute, stimulus content) analy-

sis of variance was performed collapsing over the insignificant memory-reproduction factor and excluding the first-presented nonrandomized line stimulus. The Poor-Good \times Chronic-Acute interaction was again significant, but content was not significant ($p > .05$). This result indicated that the three content scenes (trees, feeding, scolding) had the same effect on all groups. The lower size estimation evident on the first-presented stimulus, the black line, was probably due to its order in the sequence. However, the present design does not completely rule out the possibility that the simple content of the line scene did not create the differential stimulus effect. In order to examine the effect of sequence irrespective of content, a repeated measure, three-way (poor-good, chronic-acute, sequence) analysis of variance (Winer, 1962) was performed including all presented stimuli and collapsing over the insignificant memory-reproduction factor. The sequence factor was highly significant ($F = 6.02$, $df = 7/476$, $p < .01$) and indicated a gradual increase in size estimation over trials regardless of stimulus content. The sequence factor was also examined in another analysis which excluded

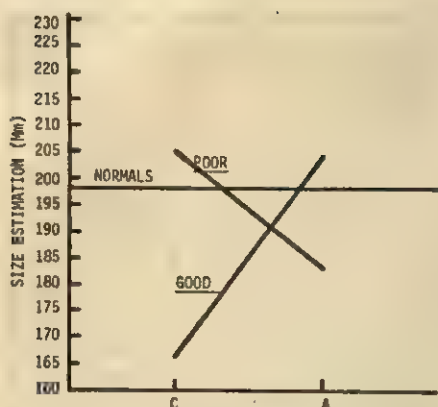


FIG. 2 THE POOR-GOOD \times CHRONIC-ACUTE INTERACTION COLLAPSED OVER CONTENT AND MEMORY-REPRODUCTION CONDITIONS.

the performance on the nonrandomized line stimulus. Sequence was again significant ($F = 4.64$, $df = 5/340$, $p < .01$). Figure 1 presents the performance of each subgroup by trials with content randomized as well as the performance of each group by content with sequence randomized. Although content was randomized in order of presentation for each group, the censure stimulus did not produce the poor-good difference as predicted by the social-censure hypothesis. There was a suggestion of the censure stimulus differentially affecting the poor-good groups in that all groups followed the same pattern of increasing or decreasing their size estimation on each scene except on the censure scene where both groups increased their size estimation and both good groups remained constant (see Figure 1). However, such a suggestion must be quite tenuous in light of not finding a significant Premorbid Adjustment \times Content interaction.

The significant Chronicity \times Premorbid Adjustment interaction is presented in Figure 2 and is due to the good underestimating in the chronic condition and overestimating in the acute condition, while the opposite occurred for the poors. The consistent relative position of each subgroup on each stimulus scene indicates that the Poor-Good \times Chronic-Acute interaction was consistent across stimuli. The significant interaction supports our main hypothesis that chronicity and premorbid do interact on a size-estimation task. The specific deviation tendencies of each group, however,

do not confirm the predictions derived from the inferences of Silverman (1964).

The significant individual group comparisons (Studentized Range Statistic, Winer, 1962) on each stimulus scene are presented in Table 2 and indicate that the chronic goods significantly differed from the acute-good and chronic-poor groups as well as normals on all stimuli. As can be seen in Figure 2, the chronic goods were the most deviant group by exhibiting the smallest size estimations, while the acute goods and chronic poors exhibited the largest size estimations. The acute-poor group was intermediate as were the normals. Since all groups including normals underestimated in relation to the standard, the specification of under- or overestimation tendencies of a particular schizophrenic subgroup was made in relation to the size-estimation performance of the normals. Poors tended to exhibit a larger size estimation than goods on all stimuli; however, the higher size estimations of the poors as compared with the goods were mainly due to the high size estimations of the chronic poors and the low size estimations of the chronic goods. There were no chronic-acute differences collapsed over premorbid adjustment

TABLE 2

SIGNIFICANT "q" TESTS BETWEEN THE SCHIZOPHRENIC SUBGROUPS OF ACUTE POORS, ACUTE GOODS, CHRONIC POORS, CHRONIC GOODS, AND NORMALS ON EACH OF THE FOUR STIMULUS SCENES

Stimulus	Subgroup	"q"
Line	AG > CG	4.98**
	CP > CG	3.74*
	N > CG	6.68**
	N > AP	3.99*
Tree	AG < CG	5.03**
	CP > CG	5.72**
	AG > AP	4.04*
	N > CG	4.16*
Feeding	AG > CG	5.30**
	CP > CG	5.27**
	N > CG	3.93*
Scolding	CP > CG	6.20**
	AG > CG	5.59**
	N > CG	4.03*

* $p < .05$.

** $p < .01$.

Note.—Significant "q" tests (k , $df = 4/68$). See note to Table 1 for explanation of abbreviations.

($p > .05$), although acutes tended to exhibit larger size estimations than chronics. The difference between chronics and acutes tended to be greater for goods than poors.

An analysis of the paranoid-nonparanoid dimension compared the paranoids and nonparanoids in each subgroup and collapsed over all subgroups. Results indicated no significant differences ($p > .05$) in size estimation between the paranoids and nonparanoids. There was a trend for the paranoids in the acute-poor and chronic-good groups (the two groups which underestimated) to display larger size estimations than the nonparanoids in these groups. This suggests that the more extreme underestimations were due to nonparanoids.

DISCUSSION

The results indicated that schizophrenic task performance was determined more by subject dimensions than stimulus conditions. Stimulus content did not exhibit any significant effects. The interaction of premorbid adjustment and chronicity was significant and remained constant across stimuli which did not allow any significant content effects. The individual comparisons indicated that the greatest between-group differences were exhibited by the large size estimation of acute goods and chronic poors as compared to the smaller size estimations of the chronic goods. The acute poors slightly underestimated as compared with normals, but were not significantly different from normals or chronic goods.

The motivational quality of maternal censure which was considered to increase the anxiety level or drive of the poors relative to the goods and normals and consequently interfere with task performance (Rodnick & Garmezy, 1957) was not supported in the present results. There was a slight tendency for censure to increase the size estimation of the poors while the goods remained relatively constant; however, this tendency did not receive any statistical support. It is possible that the present design was not able to sufficiently separate the effects of stimulus content from the effects of subject variables. The censorious cues used in the present study may not have been intense enough to increase drive and affect performance beyond the effect of the organismic variables. Also, the medi-

cated state of the present Ss may have diminished the effect of the censure stimuli (Goldstein, Acker, Crockett, & Riddle, 1966) although there was no evidence of a relationship between medication and size estimation in any of the major subject groupings.

The interaction of the subject variables of chronicity and premorbid adjustment, the significant sequence effect, and the consistency of subgroup differences across stimulus conditions lent support to the organismic model proposed by Silverman (1964). In regard to the effect of sequence, Rodnick and Garmezy (1957) suggested that in a repeated-measure design comparing different stimulus content, the censure scene should be placed last to avoid producing a carry-over effect to other stimuli. The significant sequence effect found in the present study suggests that order of presentation may produce a more potent carry-over effect than stimulus content.

Although a Chronicity \times Premorbid Adjustment interaction was found in the present results, the specific predictions concerning the performance of each subgroup were not confirmed. In fact, present subgroup performance tended to be exactly the opposite to what was predicted by Silverman. It is proposed that this difference was due to Silverman's means of inferring the scanning disposition of each subgroup, and that an interpretation of the significant findings and trends in the present data as well as a reinterpretation of previous work would allow a speculative but more consistent view of the size-estimation performance of the different schizophrenic subgroups. In inferring the tendency for the chronic goods to overestimate, Silverman relied mainly on the data of Zahn (1959) who found in comparing goods and poors that poors equaled normals and goods overestimated. Zahn considered this result due to two main factors: First, the factor of chronicity in that mean hospitalization for goods was 6.00 yr. although the sample could have ranged from very acute to very chronic; second, the factor of anxiety in that the goods who were responsible for the size overestimation were only those who were *not* on medication. In short, the overestimation could have been due to the acute goods in the sample, or the goods who were at a high-anxiety level, or both

The present results found for nonparanoids in both acute- and chronic-good groups that size estimation increased with a decrease in amount of medication, and in effect, supported Zahn's observation. Also the present acute goods tended to overestimate in comparison with normals. Therefore, Zahn's results which Silverman considered to represent the performance of chronic goods may in fact be more appropriate to acute goods. The inferences concerning the effects of chronicity on good, premorbid-adjustment schizophrenics from ad hoc analysis of Zahn's results may have been inappropriate since chronicity was not controlled and anxiety seemed to have such a determining effect. It is even possible that the medicated goods who performed similarly to normals in the Zahn study were actually a combination of the present low-estimating chronic goods and the high-estimating acute goods who when combined would appear similar to normals.

The inferences concerning the overestimation tendency of the acute poors were drawn from a study by Harris (1957). The same difficulty resides in this study as in the Zahn study. The average hospitalization of the overestimating poors was approximately 3 yr. which again could represent a range of chronicity. Inferences from the performance of this group as being representative of acute poors would be extremely hazardous. The average chronicity of 3 yr. corresponded to the upper limit of the present acute group. The performance of the Harris poors corresponds to the present chronic poors who exhibited the greatest overestimation. Since the present acute poors did not underestimate extensively from normals (especially when performance over trials was considered), the overestimation result of Harris' poors may mainly reflect an averaging of the performance of chronic poors and acute poors. Therefore, Silverman's inference of the size-estimation tendencies of acute poors may again have been premature and instead more relevant to chronic poors. The goods in the Harris study who did not differ from normals and were also considered to be acute demonstrated a slight tendency to underestimate. Due to his 3-yr. average, there possibly were more acute than chronic goods in his good group.

Their performance would, therefore, be more similar to a combination of the present chronic and acute goods with the performance level of the acute goods being more heavily weighted because of their greater number. Such a combination of the present good groups would result in a performance similar to the goods of Harris.

The present results would suggest a modification in Silverman's classification of which subgroups are characterized by over- or underestimation. The chronic goods underestimate and are most clearly different from all other groups including normals. The acute poors tend to underestimate. The chronic poors and acute goods overestimate and are not significantly different from one another.

The underestimation of all groups relative to the standard may have been due to the requirements of the present task. The size-estimation task used in previous studies (Gardner, Jackson, & Messick, 1960) required *S* to adjust circles of light with a relatively small distance between the standard and the comparison. In the present task there was a much greater distance between the *S* and the stimulus which in itself could account for smaller size estimations. Also, the content embedded in the stimuli and the drawing of a line in a field restricted by the size of the paper may have tended to decrease size estimation. Therefore, there is the possibility that although the simpler size-estimation task used in previous work has been found to be a valid measure of the extensiveness of eye movement (Gardner, Holzman, Klein, Linton, & Spence, 1959; Gardiner & Long, 1962), the distance and stimulus content factors introduced in the present study may have allowed the admittance of other factors besides that of scanning. Also, it could be questioned if the traditional psychophysical method, a variant of which was used in the present study, is capable of detecting subgroup differences in size estimation. Price and Eriksen (1966) stressed separating the judgmental process from perceptual sensitivity in analyzing schizophrenic subgroup differences. The traditional psychophysical measures, although supposedly reflecting the operation of a perceptual process such as the perception of size, may be reflecting the operation of judgment.

mental processes such as the criterion of the observer or the particular manner of using judgment categories. Magaro (1966) found that the effects of an anchor on size estimation were more a function of the use of response categories than changes in perception. Price and Eriksen (1966) found that no paranoid-nonparanoid differences could be detected in a traditional psychophysical procedure, while differences were obtained when the reliability and the S's subjective confidence in the judgmental responses were examined. The latter finding may be especially relevant to the inability of the present study to find any significant paranoid-nonparanoid differences. Although the lack of significance was probably due to the small number of paranoids in each group ($N = 6$), another interpretation could be that the present size-estimation measure was too global and insensitive to detect the type of paranoid-nonparanoid differences which may be detected with the use of more analytical dependent variables (Price & Eriksen, 1966).

REFERENCES

- BUSS, A. H., & LANG, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *Journal of Abnormal Psychology*, 1965, 70, 2-24.
- DUNHAM, R. M. Sensitivity of schizophrenics to parental censure. Unpublished doctoral dissertation, Duke University, 1959.
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1(4, Whole No. 4).
- GARDNER, R. W., JACKSON, D. N., & MESSICK, S. J. Personality organization in cognitive controls and intellectual abilities. *Psychological Issues*, 1960, 2(4, Whole No. 8).
- GARDNER, R. W., & LONG, R. I. Control, defense, and concentration effect: A study of scanning behavior. *British Journal of Psychology*, 1962, 53, 129-140.
- GOLDSTEIN, M. J., ACKER, C. W., CROCKETT, J. T., & RIDDLE, J. J. Psychophysiological reactions to films by chronic schizophrenics: I. Affects of drug status. *Journal of Abnormal Psychology*, 1966, 71, 335-344.
- HARRIS, J. G., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *Journal of Personality*, 1957, 25, 651-671.
- HOLLINGSHEAD, A. B. Two factor index of social position, Yale University, New Haven, Connecticut, 1960. (Mimeo)
- KLEIN, E. B., CICCHETTI, D., & SPOHN, H. A test of the censure-deficit model and its relation to premorbidly in the performance of schizophrenics. *Journal of Abnormal Psychology*, 1967, 72, 174-181.
- MAGARO, P. A. The effect of imagined anchors on adaptation-level. *American Journal of Psychology*, 1966, 79, 195-204.
- MAGARO, P. A. Perceptual discrimination performance of schizophrenics as a function of censure, social class and premorbid adjustment. *Journal of Abnormal Psychology*, 1967, 72, 415-420.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- PRICE, R. H., & ERIKSEN, C. W. Size constancy in schizophrenia: A reanalysis. *Journal of Abnormal Psychology*, 1966, 71, 155-160.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. *Nebraska Symposium on Motivation*, 1957, 5, 109-184.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 362-379.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- YATES, A. J. Psychological deficit. *Annual Review of Psychology*, 1966, 17, 111-144.
- ZAHN, T. P. Acquired and symbolic affective value as determinant of size estimation in schizophrenic and normal subjects. *Journal of Abnormal and Social Psychology*, 1959, 58, 39-47.

(Received November 20, 1967)

CONCEPT IDENTIFICATION BY SCHIZOPHRENIC AND NORMAL SUBJECTS AS A FUNCTION OF PROBLEM COMPLEXITY AND RELEVANCE OF SOCIAL CUES¹

VLADIMIR PISHKIN²

AND

LYLE E. BOURNE, Jr.

Veterans' Administration Hospital, Oklahoma City

University of Colorado

Schizophrenic and normal Ss ($N=230$) solved concept-identification (CI) problems varying in stimulus complexity, simultaneously with either a normal or a schizophrenic stooge providing relevant or irrelevant cues. Major results were (a) CI performance was a negative function of stimulus complexity, (b) social cues from a normal stooge had greater effects on problem solving than social cues from a schizophrenic stooge, and (c) schizophrenics demonstrated a deficit in problem solving connected primarily with the use of social cues. Comparisons of these data with expectations based on a mathematical model of CI revealed several significant discrepancies, attributable primarily to the fact that normals are more sensitive to social cues than are schizophrenics.

Pishkin and Blanchard (1963) reported a preliminary evaluation of the ability of schizophrenic and normal Ss to use social (i.e., from another person) and nonsocial (i.e., from variations in a physical stimulus) sources of information in solving concept-identification (CI) problems. Generally, their results indicated that, in addition to being poorer problem solvers, schizophrenics showed a relative inability to profit from solution-relevant information provided by another person (actually, a confederate of *E* acting as another *S*). These results, and others like them (Davis & Harrington, 1957; Johannsen, 1959), are consistent with Cameron's (1947) "social disarticulation" characterization of schizophrenic behavior. Moreover, as shown by Pishkin and Blanchard, they are describable in quantitative detail by a mathematical model of CI performance (Bourne & Restle, 1959).

The present study is, in several respects, an extension of the Pishkin-Blanchard work. First, in the earlier experiment, social cues (when they were available) were supplied by

a peer confederate (C) (i.e., schizophrenic Ss were paired with schizophrenic Cs and normal Ss with normal Cs). An issue, just as important as the difference between normal and schizophrenic ability to use social and nonsocial cues in problem solving (but not nearly as thoroughly explored), concerns the relative saliency of cues *provided by* normal and schizophrenic partners. Recent findings (Pishkin, 1966) indicate that, in general, Ss fail to use as much information from a schizophrenic as from a normal person when making perceptual judgments. The present experiment, studying the four possible dyads formed with schizophrenic and normal Ss and Cs, sought to determine whether similar performance trends are evident in a complex problem-solving task.

Second, in the preceding study all experimental problems were characterized by a single level of stimulus complexity. It is well known that stimulus complexity not only governs the difficulty of a CI task (e.g., Bourne, 1957) but also tends to interact with other significant variables (e.g., Bourne & Restle, 1959; Pishkin, 1960). The Pishkin-Blanchard results might, therefore, be severely limited in generality, with greater (lesser or reversed) differences between normal and schizophrenic performances existing in problems of other complexity levels. Accordingly, tasks of three different complexities (the

¹ The authors are indebted to Harold L. Williams for critical reading of the manuscript. This research was supported in part by Veterans' Administration Medical Research Funds.

² Requests for reprints should be directed to Vladimir Pishkin, Veterans' Administration Hospital, 921 Northeast 13th Street, Oklahoma City, Oklahoma 73104.

intermediate of which was used by Pishkin and Blanchard) were employed in the present experiment to study the possible moderating influence of this variable on normal-schizophrenic differences.

Thus, the major purpose of this work was to collect more extensive evidence on the ability of schizophrenic and normal *Ss* to solve conceptual problems of varying levels of difficulty using information provided by either or both social and nonsocial sources. In addition, an attempt was made to describe these data quantitatively within the framework of a mathematical model used by Pishkin and Blanchard. The accuracy of that description reflects one major assumption—that all sources of information, social and nonsocial, are additive—and thus serves as a minor test of the model.

METHOD

Subjects

The *Ss* were 115 schizophrenics and 115 normals. They consisted of male patients and aides from Veterans' Administration Neuropsychiatric Hospitals at Oklahoma City, Oklahoma; Little Rock, Arkansas; Tuskegee, Alabama; Tomah, Wisconsin, and Central State Hospital, Norman, Oklahoma; all patients had a current diagnosis of schizophrenic reaction, chronic-undifferentiated type. All patients (a) who were below 20 or above 45 yr. of age, (b) who had had electroshock or insulin therapy or psychosurgery, or (c) who were suffering from neurological, auditory, or visual disorders as indicated by the admissions' physical examination were not included in the sample. None of the patients were on a psychotropic drug regimen which exceeded a daily dose of 100 mg. of chlorpromazine. The schizophrenic *Ss* were first-psychiatric-admission patients with an average length of hospitalization of 48.14 days. The normal population consisted of male psychiatric aides in corresponding institutions who were also 20–45 yr. of age. All *Ss* were administered the vocabulary portion of the Shipley Institute of Living Scale prior to the experiment.

Five schizophrenic and five normal *Ss* were assigned to each of 46 experimental treatment conditions. The *Ss* were matched across all conditions for age and intelligence (Shipley vocabulary score). The schizophrenic patients were matched for length of hospitalization in each of the experimental conditions. All matching was to the extent that no significant difference was reached between any pair of groups in age, intelligence, or hospitalization.

In those conditions where a *C* was required, chronic-undifferentiated, schizophrenic patients in good contact, aides, and technicians were used. The *Cs* were dressed either in Veterans' Administration

aide or patient uniforms as appropriate and were strangers to *Ss*. Aides always wore aide uniforms and patients always wore Veterans' Administration patient uniforms. A total of 18 schizophrenic patients and 14 aides participated as *Cs*.

Procedure

All *Ss* were administered a two-choice CI problem wherein they were required to learn to match their responses with the relevant dimension of geometric patterns. The correct and incorrect responses were not so defined to *Ss*. Form of figure was the relevant dimension; that is, *Ss* were required to respond by saying "A" or "B" for squares or triangles, respectively, regardless of the other (irrelevant) characteristics of the patterns. Three levels of complexity were established by varying one (color), three (color, size, and number) or five (color, size, number, vertical position, and orientation) irrelevant dimensions of the stimuli. The patterns were presented by a male *E* on 3×5 cards mounted in a slot above a T-shaped structure. Below the stimulus slot at *S's* eye level were mounted two amber feedback lights which were positioned below the blocked letters, A and B. The feedback was provided simultaneously to *S* and *C* immediately following *S's* response. The long arm of the T structure prevented *S* and *C* from viewing each other, although both were able to view the stimulus and the feedback lights simultaneously.

Design

The design was a $5 \times 3 \times 2 \times 2$ factorial with five stimulus-social cue conditions (two of which had either stimulus or social cues only), three degrees of problem complexity, two *S* populations, and two types of *Cs*. All *Ss* performed on the task for 192 trials or until 16 consecutively correct responses were made. All patterns appeared equally often but with the restriction that the same pattern never appeared on successive trials; otherwise the order of presentation was random. In conditions where *C* participated he always responded before *S* in accordance with a sequence of responses programmed on a slip of paper. The T structure prevented *S* from viewing the *C's* program. The *S* was introduced to *C* as a patient or an aide, but was not aware that *C* was a stooge. Both *S* and *C* were given the same instructions. The *S* was informed that because *C* happened to arrive at the laboratory earlier he would always respond first and that the same feedback would be provided to *C* and *S*. A more detailed description of the task and instructions appeared in an earlier paper (Pishkin & Blanchard, 1963). The five basic conditions used in this experiment were designed in order to test the effects of availability of stimulus and social cues by combining relevant and irrelevant aspects of the two types of information. Table 1 summarizes these conditions along with the theoretical equations (developed in Pishkin & Blanchard, 1963) for predicting performance within each. As noted above, the basic CI conditions were elaborated to include three levels of task complexity. In

TABLE 1
MAIN EXPERIMENTAL CONDITIONS AND THEIR
CORRESPONDING THEORETICAL EQUATIONS

Condition	Critical solution-relevant information	Equation	N
I	st	$\theta = \frac{kR}{R + I + B}$	30
II	soc	$\theta = \frac{lS}{B + S}$	20
III	st + soc	$\theta = \frac{kR + lS}{R + I + B + S}$	60
IV	st - soc	$\theta = \frac{kR}{R + I + B + S}$	60
V	soc - st	$\theta = \frac{lS}{R + I + B + S}$	60

Note.—st=stimulus information, soc=social information, θ =learning rate, k =proportion of available stimulus information used, l =proportion of available social information used, R =number of relevant stimulus dimensions, I =number of irrelevant stimulus dimensions, B =weight of background cues, S =weight of social information.

Conditions (Cond.) II, III, IV, and V, where social information was available, either a schizophrenic (SC) or a normal (NC) stooge participated, with an equal number of schizophrenic and normal Ss paired with the two types of C. To summarize, the tasks over the five conditions were as follows:

Condition I. The Ss performed individually on a CI problem and C was not present.

Condition II. CI patterns were not available and Ss were required to guess which feedback light would go on after C had responded. The maximum number of "correct" guesses could be reached only by imitating C.

Condition III. A CI problem was presented on which C was correct on every trial. Thus, correct responses could be made by either responding to relevant stimulus dimensions or imitating C; both the stimulus and C cues were relevant.

Condition IV. This was a CI problem on which C responded randomly. Solution could only be reached by ignoring C and by identifying the relevant CI dimension.

Condition V. The C was arbitrarily correct on every trial and the pattern of guesses called correct had no relationship to the CI patterns, so that CI patterns would have to be ignored and solution could be reached only by imitating C.

RESULTS AND DISCUSSION

Analysis of variance was performed on errors to criterion and is summarized in Table 2. It indicates that the main effects of complexity, C type, and stimulus conditions were statistically significant, whereas there was no overall influence upon CI performance of S type. Consistent with previous results, it was

demonstrated that as complexity increased the number of errors also increased linearly (Pishkin, 1960). The mean number of errors when Ss, Cs, and conditions were pooled were 11.44, 25.45, and 37.82 for problems with one, three, and five irrelevant dimensions, respectively. The main effect of C showed that more errors were made when the origin of social cues was a schizophrenic stooge, as can be seen in Figure 1. This effect, coupled with the insignificant variance associated with S type, suggests that Pishkin and Blanchard's observation of an overall difference in problem solving between normals and schizophrenics is at least partly attributable to the confounding of S type and C type in their experiment. Figure 1 also indicates that the main effect of stimulus conditions is probably a result of fewer errors occurring in Cond. I, II, and III than in Cond. IV and V. This result suggests that the conflict in social and stimulus cues, which arises only in Cond. IV and V, is a significant inhibitor of problem solving.

The C type main effect is further elaborated by certain interactions. First, a significant Complexity \times C Type interaction indicates that as complexity increased from one to five irrelevant dimensions comparatively poorer performance resulted in the conditions where C was schizophrenic as compared to normal. The significant $S \times C$ Type interac-

TABLE 2
ANALYSIS OF VARIANCE OF ERRORS
TO CRITERION

Source of variance	df	F
Complexity (A)	2	8.34***
Subject type (B)	1	—
Confederate type (C)	1	14.16***
Stimulus conditions (D)	4	7.89***
A \times B	2	—
A \times C	2	3.84*
A \times D	8	—
B \times C	1	4.08**
B \times D	4	—
C \times D	4	4.05**
A \times B \times C	2	3.19*
B \times C \times D	4	—
A \times C \times D	8	—
A \times B \times D	8	—
A \times B \times C \times D	8	—
Error	172	(6236.53) ^b

* $F < 1.0$.

^b Number in parentheses = error mean square.

** $p < .05$.

*** $p < .01$.

*** $p < .001$.

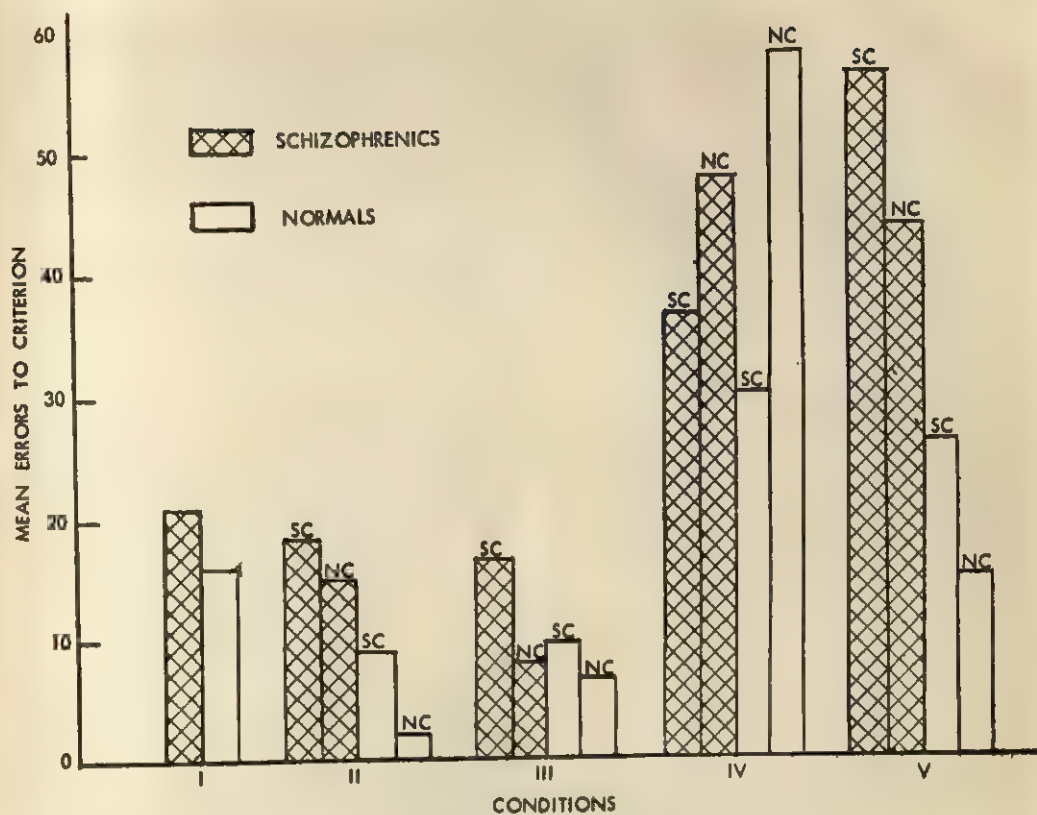


FIG. 1. Mean errors produced by schizophrenic (crosshatched) and normal (blank) Ss in the five conditions.

tion indicates, moreover, that the potency of normal Cs was significantly greater than that of schizophrenic Cs in facilitating or inhibiting CI performance of normals; schizophrenic Ss, on the other hand, demonstrated a tendency to be less differentially influenced by two types of stooges, regardless of whether they provided relevant or irrelevant information. The significant C Type \times Stimulus Condition interaction clearly demonstrates that the number of errors committed by both S populations was a function of whether C and/or stimulus cues were relevant or irrelevant to the solution of the problem; for example, as seen in Figure 1, performance in Cond. III, where both C and stimulus were relevant, was markedly superior to that of the Cond. IV where the social cues were irrelevant to the solution.

An interaction involving complexity, S, and C types emphasizes the significant trends

found in the Complexity \times C Type and S \times C Type terms. It is noteworthy that performance became progressively poorer when complexity increased, especially by schizophrenic Ss with schizophrenic Cs, whereas performance of normal Ss at the lowest level of complexity (one irrelevant dimension) with a relevant normal C was superior to all other groups.

In order to evaluate the between-group differences with both S populations and C types, a series of difference scores plus the "criterion difference" (Lindquist, 1953) required to reach statistical significance ($p < .05$) was computed for each pair of group scores, and these are presented in Table 3.

With pooled populations it is clear that there were no significant differences in errors between any pairs of combinations in Cond. I, II, and III. However, as seen in Figure 1, a significantly greater number of errors was pro-

TABLE 3
COMPARISON OF MEAN ERRORS
BETWEEN CONDITIONS

Condition	Schizophrenic S	Normal S		
A. Difference between schizophrenic C and normal C for two S populations, CD ^a = 6.32				
II	3.8	8.8*		
III	8.9*	2.6		
IV	-14.5*	-27.7*		
V	24.1*	10.9*		
B. Difference between S populations matched with different Cs, CD = 7.84				
II	10.2*	17.2*		
III	7.4	1.1		
IV	2.9	-10.3*		
V	30.5*	17.3*		
C. Difference between mean errors of schizophrenic and normal Ss, CD = 9.19				
I	4.7			
II	12.5*			
III	4.2			
IV	4.3			
V	24.1*			
D. Pooled S populations: Difference in mean errors, CD = 12.67				
	Condition			
	I	II	III	IV
II	7.2			
III	8.3	1.1		
IV	-23.9*	-31.1*	-32.2*	
V	-14.0*	-21.2*	-22.3*	9.9

* Criterion difference.

* $p < .05$.

duced in Cond. IV and V as compared to Cond. I, II, and III (Table 3, Section D).

Section A of Table 3 represents differences in mean errors for schizophrenic and normal Ss between dyads performing with schizophrenic and normal Cs. A positive score indicates a greater number of errors with schizophrenic Cs. Significant differences for Cond. III, IV, and V indicate that the influence of schizophrenic Cs on performance of schizophrenic Ss was significantly less than that of the normal Cs. With normal Ss it is also clear that schizophrenic Cs had a less facilitating influence upon performance than normal Cs.

However, it is noteworthy that in Cond. III, where both stimulus and social cues were relevant, there were no differences in the performances of normal Ss as a function of C type; this finding is most likely due to the fact that the problem was sufficiently simple for normal Ss that it masked the effect of C type.

Section B of Table 3 represents differences between the two S populations due to C type. A positive score in this section indicates a greater number of errors produced by schizophrenic Ss than by normal Ss. It is apparent that when NCs were used in Cond. II and V, a significantly greater number of errors was produced by schizophrenic Ss than by normal Ss. These data are consistent with other findings that schizophrenics are not able to gain relevant information from a social source to the same degree as normal Ss. Poorer performance by normals was evident in Cond. IV, however, where the function of C was to interfere with solution.

Section C of Table 3 shows differences between the two S populations for each condition. The only significant differences appeared in Cond. II and V, again demonstrating deficiency of schizophrenic Ss in using social information, regardless of the source. The fact that no differences were found between the two populations in Cond. III is probably attributable to the simplicity of the problem. In Cond. IV the interference effect of C for normal Ss might have been sufficient to overcome their typical advantage over schizophrenics, since normal Cs negatively influenced performance of normal Ss to a greater degree than that of schizophrenic Ss.

Theoretical Analysis

Calibration of theoretical parameters in Table 1 was accomplished using Equations 1, 2, and 5 (derived in Pishkin & Blanchard, 1963). Calibration data were taken from two subgroups of Cond. I (one and five irrelevant dimensions), Cond. II, and one subgroup of Cond. V (five irrelevant dimensions). These parameters were derived by use of simultaneous equations and applied in predicting performances for the remaining groups. The calibrated values are summarized in Table 4. Consistent trends are apparent: The rate of

stimulus-cue usage, the rate of social-cue usage (from a schizophrenic or normal source), as well as the weights of social-stooge cues—all of which are quantitative parameters, estimated from the data and required for an accurate fit between model and data—were smaller for schizophrenic than normal Ss. In addition, it is clear that the value of normal social cues was approximately twice as great as the value of schizophrenic social cues for both the normal and schizophrenic S populations.

Table 5 shows the predicted mean errors for conditions which were not included in the calibration of constants. It is apparent that the two predictions in conditions with stimulus cues only were adequate. Also the prediction of combined and redundant stimulus and social cues (Cond. III) was fairly close, at

TABLE 4

CALIBRATION OF THEORETICAL PARAMETERS

Group	k	B	l_s	S_s	l_n	S_n
Normals	.45	1.43	.26	1.55	.39	2.65
Schizophrenics	.37	1.30	.18	.93	.15	1.85

Note.— k =proportion of available stimulus information used; B =weight of background cues; l_s =proportion of cues from schizophrenic stooge used by S ; l_n =proportion of cues from normal stooge used by S ; S_s =weight of schizophrenic C; S_n =weight of normal C.

least for normal Ss. However, there were severe discrepancies when stimulus and social information conflicted (Cond. IV and V), a result which reflects primarily the fact that normal Ss were significantly more sensitive than schizophrenics to social cues. The assumption of equivalent relative sensitivity made by the model is simply inappropriate for a description of these data.

TABLE 5

OUTCOME OF THEORETICAL ANALYSIS TO CRITERION

Condition and no. irrelevant dimension	Stooge ^a	Group					
		Schizophrenic			Normal		
		θ Predicted	Error		θ Predicted	Error	
			Predicted	Obtained		Predicted	Obtained
I. Stimulus only							
One		(.111) ^b	—	(11.4)	(.132) ^b	—	(8.8)
Three		.07	19.8	(19.8)	.083	16.2	(16.2)
Five		(.05) ^b	—	(31.6)	(.061) ^b	—	(23.6)
II. Social only							
None	S	(.074) ^b	—	(18.6)	(.137) ^b	—	(8.4)
	N	(.09) ^b	—	(14.8)	(.255) ^b	—	(3.6)
III. Stimulus + social							
One	S	.13	9.0	(4.8)	.171	6.0	(5.8)
	N	.13	9.0	(3.8)	.24	3.8	(3.4)
Three	S	.09	14.6	(14.2)	.122	9.6	(7.4)
	N	.091	14.3	(6.0)	.18	5.5	(6.8)
Five	S	.06	24.9	(31.8)	.095	13.7	(15.4)
	N	.071	19.0	(14.2)	.147	7.6	(10.4)
IV. Stimulus-social							
One	S	.087	15.6	(14.6)	.09	14.6	(12.8)
	N	.072	19.0	(16.0)	.074	19.0	(24.6)
Three	S	.059	26.5	(36.2)	.06	24.8	(32.2)
	N	.050	31.5	(41.4)	.055	28.0	(51.8)
Five	S	.045	37.0	(49.4)	.05	31.5	(46.4)
	N	.04	41.6	(86.2)	.044	37.0	(96.4)
V. Social-stimulus							
One	S	.04	41.0	(21.4)	(.082) ^b	—	(16.4)
	N	.054	27.0	(10.2)	(.171) ^b	—	(6.2)
Three	S	.03	60.0	(52.0)	.057	27.0	(24.2)
	N	.04	41.6	(30.4)	.127	9.3	(17.8)
Five	S	(.02) ^b	—	(98.2)	.045	37.0	(37.6)
	N	(.031) ^b	—	(56.8)	.10	12.7	(21.6)

^a Confederate type; S=schizophrenic, N=normal.

^b Estimated from data.

Some general conclusions made from this theoretical analysis include (a) normals used more of the relevant stimulus information than schizophrenics, (b) since the *B* parameter represented irrelevant, background stimulation arising from sources other than the stimulus pattern, it is apparent that normals were more sensitive to background cues than were schizophrenic Ss, (c) normals were more sensitive to social cues than schizophrenics whether those cues came from schizophrenic or normal Cs, (d) both normals and schizophrenics used more cues from a normal C than from a schizophrenic C, (e) the performance of normals working with both stimulus and social cues followed theoretical predictions adequately, thus giving some support to the additivity of cues assumption of the model, though the same conclusion obviously does not hold for schizophrenics, (f) discrepancies between model and data resulted largely from the fact that schizophrenics did not use cues from normal Cs as effectively as predicted, and (g) the model assumes that schizophrenic Ss would use C (normal or schizophrenic) in the same way that the normal S used C. The significant interaction $S \times C$ type indicates that this particular assumption is untenable. Present findings clearly demonstrate schizophrenics' deficit in processing of information from a social source, especially when such source is a peer, schizophrenic patient.

REFERENCES

- BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *Journal of Experimental Psychology*, 1957, **54**, 201-207.
- BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review*, 1959, **66**, 278-296.
- CAMERON, N. *The psychology of behavior disorders: A biosocial interpretation*. Boston: Houghton Mifflin, 1947.
- DAVIS, R. H., & HARRINGTON, R. W. The effect of stimulus class on the problem-solving behavior of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1957, **54**, 126-128.
- JOHANNSEN, W. J. Responsiveness of schizophrenics and normals to social and non-social feedback. Unpublished doctoral dissertation, University of Wisconsin, 1959.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology*, 1960, **59**, 371-378.
- PISHKIN, V. Perceptual judgment of schizophrenics and normals as a function of social cues and symbolic stimuli. *Journal of Clinical Psychology*, 1966, **22**, 3-10.
- PISHKIN, V., & BLANCHARD, R. J. Stimulus and social cues in concept identification of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1963, **67**, 454-463.

(Received November 20, 1967)

EFFECT OF PERCEPTUAL ISOLATION AND AROUSAL ON ANXIETY, COMPLEXITY PREFERENCE, AND NOVELTY PREFERENCE IN PSYCHOPATHIC AND NEUROTIC DELINQUENTS¹

GEORGE J. SKRZYPEK²

University of Illinois

Psychopathic and neurotic delinquents were subjected to perceptual isolation (PI) and arousal (A) and were given pre- and postmeasures of novelty, complexity preference, and anxiety. The results showed that (a) there was a significant negative correlation between novelty preference and anxiety level and between complexity preference and anxiety level; (b) the psychopathic delinquents had significantly lower pretest anxiety and significantly higher novelty and complexity preference scores; (c) after PI psychopathic delinquents increased their complexity preference scores significantly more; and (d) after A psychopathic delinquents did not shift scores, while neurotics significantly increased anxiety scores and decreased complexity preference scores.

This investigation was concerned with the preference behavior of psychopathic and neurotic delinquents for novelty and complexity of stimuli under experimental conditions of perceptual isolation and arousal. Its principal aim was to test the validity of a theory concerning the etiology of psychopathy proposed by Quay (1965) which holds that the psychopathic personality represents pathological stimulus seeking. Quay (1965) states: "We are suggesting that the level and variability of sensory inputs which are necessary for the maintenance of pleasant affect are much greater for the psychopath than for the ordinary individual [p. 181]." In order to account for this pathological need for sensory input, Quay suggests two possibilities. The first is that the psychopath's basal reactivity to stimulation is lower than normal, so that

more sensory input is needed to produce efficient and pleasurable cortical functioning. The second is that there may be a more rapid adaptation to stimulation which causes the need for stimulus variation to occur more rapidly and with greater intensity.

Quay (1965) concludes that:

What we have suggested is that the psychopath, either due to a lessened basal reactivity or an increased rate of adaptation, quite frequently finds himself in a condition of stimulus deprivation. Since this condition is affectively unpleasant he is motivated to change this affective state by the seeking of stimulation. In a highly organized environment such as that in which modern man resides this seeking of either added intensity or added variability of stimulation may on occasion involve transgressions of both law and moral code. It is these transgressions, motivated by the search for added or more varied stimulus input, which most frequently bring the psychopath to public attention [p. 182].

¹The present paper is based on a dissertation submitted in partial fulfillment of the requirements for a PhD in psychology at the University of Illinois. This study was made possible by the kindness and cooperation of R. E. Gerard and R. B. Levinson at the National Training School for Boys, Washington, D. C. The author wishes to give grateful thanks to H. C. Quay and L. P. Ullmann for their helpful suggestions and continuing support throughout the planning and execution stages of this study. The author also wishes to thank G. L. Paul, J. E. McGrath, J. S. Wiggins, and M. M. Ohlsen for their helpful assistance.

²Request for reprints should be sent to the author, United States Army Hospital, United States Military Academy, West Point, New York 10996.

Another major aim of this investigation was to test the accuracy of a frequently made clinical observation concerning a characteristic which distinguishes the psychopathic from the neurotic personality. This observation is that psychopaths are persons suffering from affective impoverishment; that is, psychopaths do not become anxious in situations which are highly anxiety eliciting for non-psychopathic individuals. Preference for novel and complex stimuli and an arousal condition therefore were variables in this study, so that it was possible to explore the relationship be-

tween anxiety and complexity- and novelty-preference behavior. Several studies (Berlyne & Lewis, 1963; McReynolds & Bryan, 1956; O'Leary, 1965) suggest that there is a negative relationship between preference for complexity or novelty and anxiety. However, Barron (1953) found that complexity correlated .37 with the *Pd* scale of the MMPI, and Haywood (1962) found no significant differences among three groups of low, moderate, and high anxiety Ss on choice of novel stimuli. It was hoped that the present study would clear up these discrepant results.

Brief plan of experiment. This investigation used a group of psychopathic and neurotic juvenile delinquent males. Neurotic delinquents were included because they constitute a group which theoretically should be on the opposite extreme of an anxiety and impulse control continuum from psychopaths, and because they could be theoretically characterized as "pathological stimulus avoiders." They thus provided a marked contrast to the psychopaths. Both groups were administered a questionnaire-type anxiety scale, a preference for complex stimuli test, and a preference for novel stimuli test. Following this, both groups were randomly assigned to one of three experimental conditions. The first condition consisted of a brief perceptual isolation experience, the second was a relatively intense arousal condition, and the third was a "neutral" control condition. Immediately following exposure to these conditions, all Ss were administered a posttest of anxiety, complexity preference, and novelty preference.

It was hypothesized that the psychopathic delinquents would be relatively unaffected by the arousal condition, but would show significant increases in novelty and complexity preference as a result of perceptual isolation, while the neurotic delinquents would react in the opposite way. Furthermore, the pretest novelty and complexity scores were expected to be significantly higher for the psychopathic than for the neurotic delinquents. (For a complete statement of predictions see Skrzypek, 1967.)

METHOD

From a sample of juvenile delinquent boys at the National Training School for Boys (NTS) in Wash-

ington, D. C., a group of psychopathic and neurotic delinquents were selected on the basis of their scores on the psychopathy and neuroticism subscales of a behavior rating checklist developed by Quay and his collaborators (Peterson, Quay, & Cameron, 1959; Peterson, Quay, & Tiffany, 1961; Quay, 1964). These authors identified by factor-analytic methods four personality dimensions related to juvenile delinquency. The two dimensions relevant to the present investigation were "unsocialized-psychopathic" and "disturbed-neurotic." The psychopathic-delinquency factor was interpreted as reflecting overt, aggressive, amoral acting-out behavior, coupled with impulsivity, a rebellious attitude toward authority and discipline, and a relative freedom from family and other interpersonal ties. The neurotic factor "reflects a dimension of withdrawal, escape, anxiety and repression [Quay, 1964]."

Quay's (1964) behavior rating checklist has been shown to measure both these factors. It is comprised of a 46-item checklist drawn from Peterson's (1961) Problem Checklist. The behaviors on this checklist are easily observable and are rated with adequate reliability by cottage personnel and teachers. Quay et al. (1966) report a correlation of .67 between parents' ratings on the neuroticism subscale, and .78 between parents' ratings on the psychopathy subscale.

The mean for the psychopathy scale of the behavior rating, calculated on an *N* of 360 delinquents from NTS, was 7.973, *SD* = 4.98. The mean for the neuroticism scale, calculated on the same *N*, was 7.932, *SD* = 4.19. The Ss were selected from a pool of 200 boys who were in the institution at time of testing. All remaining Ss who attained a psychopathy scale score of greater than 8 and a neuroticism scale score of less than 8 were placed in the psychopathic group. All Ss who attained a neuroticism scale score of greater than 8 and a psychopathy scale score of less than 8 were placed in the neurotic group. There were a total of 33 Ss who met this criterion for psychopathy, and a total of 33 who met the criterion for neuroticism.

Dependent variables. There were three main dependent measures used in the present study: (a) The Anxiety Differential (Husek & Alexander, 1963), an 18-item test in the form of a semantic differential scale. (b) A measure of the degree of preference for complex versus simple random shapes (Munsinger & Kessen, 1964), comprising two different sets (for pre- and posttest) of 28 pairs of pictures of random shapes. Each pair contains one "complex" and one "simple" shape, complexity and simplicity being defined by the number of independent turns that each shape possesses. (c) A measure of the degree of preference for "novel" versus "familiar" pictures (Sieveking, 1965). This test consists of two different sets of 12 pairs of pictures, one picture in each pair being an unusual or incongruous picture, the other being a familiar one.

Procedure. The Ss were run through the experimental procedures individually. Upon entering the experimental room *S* was introduced to *E* and given

a brief sketch of what he would be required to do in the experiment. The *E* then informed *S* in the perceptual isolation and control conditions that if he participated in the study and cooperated satisfactorily, he would receive \$2. The *Ss* in the arousal condition were told that they would receive a *chance* to earn up to \$2 if they participated and cooperated by following instructions. The *S* was then asked whether he desired to participate in the experiment. Only one *S*, who was classified as a psychopath, refused to participate. Prior to seeing *Ss*, the entire potential *S* list was shuffled and coded as to psychopathic and neurotic classification by another psychologist working at NTS. Therefore, *E* was blind as to *S* classification.

The *Ss* were first administered the Anxiety Differential test. To insure that *Ss* understood how to fill out the test properly, three examples were worked by *S* with *E*'s help, and *E* read all the words on the test to *S*, in case *S*'s reading ability was too poor for the test. Next the precomplexity-preference and the prenovelty-preference tests were administered. The order of administration of these two measures was counterbalanced. The three experimental conditions, perceptual isolation (PI), arousal (A) and control, followed. The psychopathic and neurotic delinquents were randomly assigned to these conditions, with the restriction that there would be 11 *Ss* in each condition.

The PI condition consisted of requiring *S* to lie quietly on a bed with his head on a foam rubber pillow. Each *S* in this condition wore thick cotton gloves, translucent goggles, and earphones through which white noise of a sufficient intensity to mask all outside sounds was fed. The *Ss* in PI were informed that an electrical timer and white-noise generator were actually equipment for recording every movement they made. They were also told that other boys like them did very well in relaxing, and that *E* wanted to see how well they could do. The *Ss* were not informed of how long they were to remain in this condition. After administering the instructions, *E* turned off the lights, turned on the tape recorder which introduced the white noise through the earphones, and left the room. He quietly looked in upon each *S* at 2-min. intervals to insure that *S* was following instructions. Each *S* remained exactly 40 min. in perceptual isolation.

The *Ss* in the control condition were told that while *E* was engrossed in working on some numbers, they were to keep busy tracking a moving target on the pursuit rotor. The *E* then demonstrated the use of the pursuit rotor, and explained the electric timers to *S*. He then set *S* to tracking, and pretended to busy himself with adding up columns of figures. Each *S* completed eight blocks of 3 min. per block of tracking, with 2-min. rest periods in between blocks. The total time for the control condition was thus 40 min.

The arousal condition consisted of *Ss* making successively more difficult auditory discriminations between the loudness of two tones. Pure tones of 1,000 cps were used. The intensity of the constant tone

was 40 db. SPL. The comparison tone ranged in intensity from 53 to 40 db. SPL. At the beginning of the actual test trials, the intensity difference between the constant and comparison tone was 13 db. SPL. From there this difference decreased at a rate of 1 db. per step, with 2 trials given to *S* per step. There were a total of 32 trials given to each *S*, with an intertrial interval of 10 sec. In the last 6 trials the intensity of the constant and comparison tone was exactly equal. Two seconds before each trial a 600 cps warning tone, set at a constant intensity of 35 db. SPL, was presented. The purpose of this tone was to make *S* aware that the comparison and constant tones were about to be presented. The order of presentation of the constant and comparison tone was randomized throughout all 32 trials. Whenever *S* made a correct discrimination, he was given a poker chip which represented 10¢. There was a total of 20 poker chips available. Whenever an incorrect discrimination was made, *S* was first administered a 1-sec. burst of white noise set at an intensity level of 60 db. SPL, and then 1 poker chip was taken away from the total which he had amassed up to that point.

The *Ss* in this condition were instructed that the discrimination task was a test of how well they could hear. They were led to believe that it was a very easy test, and that most boys like them had performed almost perfectly on it in the past. To make the discriminations, *Ss* were instructed to say "one" if the first tone was louder than the second, and "two" if the second tone was louder than the first. Six practice trials were administered prior to the actual test trials, in order to insure that *Ss* clearly understood what they were supposed to do. These practice trials involved extremely easy discriminations, which all *Ss* who possessed normal hearing were expected to make correctly. After explaining the contingencies applying to correct and incorrect discriminations, the test trials were initiated. In order to eliminate possible cues given by *E*'s facial expressions or movements, *Ss* were required to wear translucent goggles during the test trials. Since the arousal condition took only an average of 10 min., 30 min. of tracking on the pursuit rotor was given to *Ss* in this condition after administration of the pretest and prior to the discrimination-arousal condition. This made the time interval elapsing between pre- and posttests equal to that of the other two conditions.

The novelty- and stimulus-complexity-preference measures were readministered immediately after the three experimental treatments. Again, the order of presentation was counterbalanced for each group. The posttest Anxiety Differential was administered last. (See Skrzypek, 1967, for complete procedure and instructions.)

RESULTS

Scoring. A scoring method was employed to determine the frequency with which each *S* chose the more variable or complex random

TABLE 1
MEANS AND STANDARD DEVIATIONS OF DEPENDENT VARIABLES

Ss	Variable					
	AD pretest	AD posttest	RS pretest	RS posttest	NP pretest	NP posttest
Psychopathic delinquents						
Perceptual isolation						
<i>M</i>	39.18	39.36	14.90	19.90	8.27	9.09
<i>SD</i>	8.795	10.259	6.378	4.700	2.686	2.300
Arousal						
<i>M</i>	38.09	38.09	15.72	14.28	8.00	7.81
<i>SD</i>	8.904	7.867	6.435	4.729	2.529	2.561
Control						
<i>M</i>	40.00	40.54	17.27	18.00	9.36	8.90
<i>SD</i>	8.933	7.763	6.619	6.496	2.292	2.343
Neurotic delinquents						
Perceptual isolation						
<i>M</i>	96.18	96.90	10.72	12.90	4.45	4.54
<i>SD</i>	18.334	18.571	5.497	4.570	2.339	2.423
Arousal						
<i>M</i>	93.90	111.54	13.18	8.90	4.09	3.63
<i>SD</i>	17.563	17.084	6.438	5.575	2.385	2.292
Control						
<i>M</i>	104.00	102.63	14.27	15.63	4.27	4.54
<i>SD</i>	18.681	18.554	5.640	6.888	2.649	2.544

Note.—AD = Anxiety Differential; RS = random shapes; NP = novelty preference.

shape. Each of the 28 choices in the pre- and posttests of preference was given a score of 1 if *S* chose the stimulus having greater variability and a score of 0 if he chose the stimulus with less variability. A similar scoring procedure was employed to determine the frequency of choice of the novel or incongruous picture versus the familiar one. The Anxiety Differential test was scored according to the key given by Husek and Alexander (1963). Pre- and posttest means and standard deviations of the three dependent variables for the psychopathic and neurotic delinquents under all three conditions are presented in Table 1.

Analysis of Anxiety Differential scores.

Anxiety Differential scores were analyzed by a $3 \times 2 \times 2$ repeated measures design analysis of variance (Winer, 1962). This analysis involved three levels of experimental condition, two levels of *S* classification (psychopathic and neurotic) and two testing times (pre- and posttest). The main effect of treatments along was nonsignificant. The main effect of *S* classification had a very significant affect, psychopathic delinquents as a group scoring low on the Anxiety Differential, and neurotic delinquents scoring very high. The interaction of treatments with classification was not significant. The main

effect of pre- and posttesting, and the interactions of pre- and posttesting with classification and with treatments were all significant, as was the three-way interaction, which qualifies the other two significant interactions.

The large increase in the Anxiety Differential by neurotic delinquents from pre- to posttest under the arousal condition was the factor responsible for all three significant interactions. Components of the sums of squares were calculated for the B(Classification) \times C (Testing) interaction under each of the three treatments. This resulted in mean squares and *F* ratios for each of these three interactions. The B \times C interaction under the arousal condition was highly significant ($F = 37.62$), while the *F* ratio for the B \times C interaction under the control and perceptual isolation condition was less than one in both cases. The *t* tests for correlated means were also computed on the pre- and postdifferences for psychopaths and neurotics under all three treatments. The only significant difference was for neurotics under the arousal treatment ($t = 8.93$, $df = 10$).

Two one-way analyses of variance were computed on only the pretest scores for the psychopathic and neurotic groups in order to ascertain whether there were any significant

differences among those scores within either group. There were no significant differences among the pretest Anxiety Differential scores *within* the three psychopathic or three neurotic delinquent groups. A two-way analysis of variance (Treatments \times S Classification) was also calculated on the pretest scores in order to ascertain whether there was a significant overall difference between the pretest scores for the psychopathic *versus* the neurotic groups. The main effect of S classification was highly significant, the main effect level means indicating that the pretest scores were low for psychopaths and high for neurotics.

The following conclusions were drawn from the above analyses of Anxiety Differential scores: (a) As was predicted, there were no significant differences between pre- and posttest for any of the psychopathic delinquent groups. (b) The only significant difference between pre- and posttest for the neurotic delinquents occurred under the arousal condition. Consistent with clinical theory, the neurotics significantly increased their scores from pre- to posttest. (c) Random assignment to experimental conditions was shown in that there were no significant differences between the pretest scores *within* the three neurotic or the three psychopathic delinquent groups. (d) As hypothesized, there was a significant overall difference between the pretest scores of the psychopathic delinquents and the pretest scores of the neurotic delinquents, with the neurotics having the higher scores.

Analysis of random shapes scores. Complexity-preference test scores were analyzed in exactly the same manner as were the Anxiety Differential data. The main effect of treatments alone and of testing alone did not reach significance. The main effect of S classification was again highly significant. The interaction of treatments with classification ($A \times B$) was not significant. The within-Ss interactions $B \times C$ and $C \times A$ (Classification \times Testing and Testing \times Treatments) were both significant. However, the $A \times B \times C$ interaction failed to reach the .05 level of significance ($F = 2.18$, $F = 3.15$ needed for .05 level). It was close enough to significance, however, that a closer look at its components of sums of squares seemed justified. The sums of squares, mean squares, and F ratios for

the $B \times C$ interaction were computed for each level of A (treatments). The F ratio for the $B \times C$ interaction under the perceptual isolation treatment reached the .05 level of significance ($F = 4.09$). This was also true of the $B \times C$ interaction under the arousal condition ($F = 4.15$). However, the $B \times C$ interaction under the control condition was not significant ($F = .204$).

The significant F ratio for the $B \times C$ interaction under the perceptual isolation condition means that the *increase* in preference for complexity from pre- to posttest was significantly greater for psychopathic than for neurotic delinquents. The significant $B \times C$ interaction under the arousal condition signifies that the decrease in preference for complexity from pre- to posttest was significantly greater for neurotics than for psychopaths. In order to ascertain whether the psychopaths significantly decreased from pre- to posttest under the arousal condition, a t test for correlated means was computed on the difference between their pre- and posttest means. This difference fell just short of the .05 level. The 3×2 analysis of variance (Treatments \times Classification) on pretest scores alone showed that there was an overall significant difference between the pretest scores for psychopaths compared with neurotics ($F = 4.52$), with psychopathic delinquents having higher scores than neurotic delinquents. The one-way analyses of variance on pretest scores showed that there were no significant differences among the pretest scores within the three psychopathic or the three neurotic delinquent groups.

The t tests for correlated means were computed on the pre- and posttest differences and showed that (a) the difference between pre- and posttest scores under perceptual isolation was significant ($t = 6.65$) for psychopathic and for neurotic ($t = 2.14$) delinquents; (b) the difference between pre- and posttest scores under arousal was significant for neurotics ($t = 5.87$) but short of significance for psychopaths ($t = 1.80$); and (c) the differences between pre- and posttest scores under the control condition were not significant for psychopaths ($t = .48$) or for neurotics ($t = .15$).

The following conclusions may be drawn

from the above analyses of the random shapes data:

1. Both psychopathic and neurotic delinquents increased their scores significantly from pre- to posttest under perceptual isolation. As would be predicted by Quay (1965) the increase for psychopaths was significantly greater than the increase for neurotics.

2. As expected, neurotics decreased their scores significantly from pre- to posttest under the arousal condition. The decrease for psychopaths was not significant, which is consistent with Quay's (1965) theory.

3. Random assignment to experimental conditions was shown in that there were no significant differences between the pretest scores within either the psychopathic or neurotic delinquent groups.

4. The differences from pre- to posttest under the control conditions were not significant within either the psychopathic or neurotic delinquents, indicating the adequacy of the control task.

5. As predicted, there was a significant overall difference between the pretest scores of psychopaths compared with neurotics, with psychopaths having the higher pretest scores.

Analysis of novelty preference scores. For novelty preference, the only F ratio in the three-way analysis of variance which reached significance was for the main effect of S classification. The $A \times B \times C$ interaction failed to reach significance. Nonetheless, the $B \times C$ interaction mean squares and F ratios were computed under each level of A . None of these F ratios approached significance. The t tests for correlated means were computed on the pre- and posttest differences for both neurotic and psychopathic delinquents. The only significant difference ($t = 2.32$) occurred for psychopaths under the perceptual isolation condition.

Two one-way analyses of variance were computed on the pretest scores separately for the neurotics and psychopathic groups. A two-way analysis of variance on pretest scores was also computed. Neither of the F ratios for the one-way analyses reach significance, indicating that there were no overall

significant differences among the pretest scores within either the psychopathic or neurotic groups. The main effect of S classification in the two-way analysis was highly significant, showing that the pretest scores for psychopaths were significantly different from those for neurotics. The level means for main effects indicated that psychopaths had higher pretest scores than neurotics.

The conclusions which may be drawn from the above analyses of novelty preference scores were as follows: (a) Novelty preference scores did not change significantly from pre- to posttest, with the single exception of psychopaths under perceptual isolation, who increased their preference for novelty significantly. (b) There were no significant differences among pretest scores within either the three neurotic or the three psychopathic delinquent groups, indicating that random assignment was achieved. (c) As was predicted, there was a significant overall difference between the pretest scores for psychopaths when compared with neurotics, the psychopaths having the higher scores.

Analysis of subject variables. A 3×2 analysis of variance (Treatments $\times S$ Classification) computed on age, IQ, length of stay in the institution, and number of school grades completed showed that these variables were randomly distributed in both the psychopathic and neurotic delinquent groups over the three experimental conditions. Means and standard deviations for these scores may be found in Skrzypek (1967, Table 1).

Correlations were computed between the pretest scores of the Anxiety Differential and the pretest scores of random shapes, novelty preference, and the behavior ratings. They appear in Table 2. All these correlations were significant well beyond the .01 level. The correlation between pretest scores for novelty preference and random shapes pretest scores was .805 for the psychopathic delinquents ($N = 33$) and .792 for the neurotic delinquents ($N = 33$). The substantial correlations of $-.840$ between pretest anxiety scores and the psychopathy behavior rating scale serve to validate further Quay's (1964) behavior rating scales.

Since there was a very high correlation between both the neuroticism and psychopathy

subscales of Quay's behavior checklist and the anxiety scores, the question can be raised whether anxiety is the crucial variable, and the classification as to psychopath and neurotic does not contribute significantly to the variance. With this in mind, an analysis of covariance was performed on the random shapes data. Using pretest anxiety scores as the covariate, an F ratio of 3.28 was obtained, which is significant at the .05 level. This indicates that classification still contributed significantly to the variance with anxiety scores held constant.

DISCUSSION

The overall tenor of the results was confirmatory of Quay's (1965) theory of psychopathy. The immediately following material will discuss the data in the order with which the major hypotheses were presented in Skrzypek (1967).

1. There was a high negative correlation between both preference for novelty and for complexity and the pretest scores of the Anxiety Differential. To the extent that these three measures possess validity, the present data indicate that the more anxious a person is, the less likely he is to choose novel versus familiar or complex versus simple stimuli. This is consistent with the studies of Berlyne and Lewis (1963), O'Leary (1965), and McReynolds and Bryan (1956).

2. Psychopathic delinquents had significantly lower pretest scores on the Anxiety Differential and significantly higher pretest scores on the complexity- and novelty-preference measures than neurotic delinquents. The fact that psychopaths attained lower Anxiety Differential pretest scores than neurotics serves to substantiate clinical observations and research such as that by Lykken (1957) and Hare (1965). Quay's (1965) hypothesis that the psychopath is a pathological stimulus seeker is also supported by the present results, which showed that the psychopathic delinquents had significantly higher pretest scores on both preference for complex stimuli and novelty than the neurotic delinquents.

3. Under perceptual isolation, the increase from pre- to posttest of complexity preference

TABLE 2
CORRELATIONS BETWEEN PRETEST SCORES OF THE ANXIETY DIFFERENTIAL AND RANDOM SHAPES, NOVELTY PREFERENCE, AND BEHAVIOR RATINGS

Ss	Random shape	Novelty preference	Behavior rating	
			BR-P	BR-N
Psychopaths				
Anxiety Differential	-.844	-.656	-.840	
Neurotics				
Anxiety Differential	-.807	-.868		.699
Total Ss				
Anxiety Differential	-.550	-.838	-.858	.833

Note.—BR-P = psychopathy subscale; BR-N = neuroticism subscale; $N = 33$ in each group.
 $p < .01$.

was significantly greater for psychopaths than for neurotics. This suggests that psychopaths were affected to a greater degree by the perceptual isolation procedure than were the neurotics. Perceptual isolation can be conceptualized as having produced information deprivation to the central nervous system, and this in all likelihood functioned as a motivational variable which caused the Ss to seek increasing amounts of stimulus variability. This motivational effect was stronger for psychopaths than for neurotics.

4. The decrease from pre- to posttest under the arousal condition on complexity preference was significantly greater for neurotics than for psychopathic delinquents. These results indicate that psychopaths were not affected by the arousal manipulation. Their average pre- and posttest anxiety scores were identical, and there were no significant differences between their pre- and posttest scores on preference for complexity and novelty. That neurotics were affected by the arousal treatment was indicated by the significant increase in their anxiety posttest score and the significant decrease in their complexity preference behavior. These results also support Quay's contention that the psychopath possesses a greater than average need for increases in the intensity level of stimulation.

5. As would be predicted from Quay's (1965) theory, the psychopaths in perceptual isolation received the highest of all posttest scores on complexity and novelty preference, while neurotics in the arousal condition

achieved the lowest of all scores on these measures.

Although there was a significant difference between psychopaths and neurotics in their pre- and posttest novelty-preference scores, with only one exception these scores did not shift significantly from pre- to posttesting within either group as a result of the experimental manipulations. The failure of the novelty-preference scores to shift significantly may be due to a tendency on the part of the Ss to be consistent in their choices of either the novel or familiar stimuli. Unlike the random shapes, the novel stimuli are obviously different from the familiar stimuli. Once an S commits himself to choosing either type of stimuli, he is likely to continue to make the same type of choices in an effort to make himself appear consistent. That this is a plausible hypothesis was shown by the fact that out of the total number of Ss (66), 21 had identical scores on the pre- and posttest of novelty preference and 35 shifted by only 1 point.

The fact that only a relatively brief period of perceptual isolation was sufficient to produce significant changes in complexity- and novelty-preference behavior is encouraging for future research in this area. Previous studies using sensory deprivation techniques usually involved a large number of S hours (up to 3 full days) and extensive and expensive laboratory equipment, a fact which is certainly prohibitive to most researchers. The condition used in this study involved only 40 min. of individual S time and only a limited amount of relatively inexpensive apparatus. This technique is well within the budget and time limitations of most researchers.

In conclusion, the data reported in the present investigation lend strong support to Quay's (1965) theory of psychopathy as pathological stimulus seeking. The results suggest a description of the psychopath as an individual who lacks tolerance for even short periods of inactivity and sameness, and who is highly motivated to seek added intensity and variability of stimulation after experiencing such sensory restriction. Furthermore, the psychopath seems to find himself in a state of stimulus deprivation even in the

everyday environment, as shown by the high pretest scores on novelty and complexity seeking of psychopathic delinquents compared with neurotic delinquents.

The strength of these results might be increased by the use of a normal group. However, normative data on the Anxiety Differential for normal Ss is available. Alexander³ tested 759 incoming college freshmen and found the mean of these cases to be 56.00, $SD = 10.91$. As would be predicted by clinical theory, this mean falls about midway between the combined pretest scores of psychopaths (39.09, $SD = 8.63$) and of neurotics (98.03, $SD = 18.15$).

Quay (1965) suggested that the psychopath is a pathological stimulus seeker because of either a lessened basal reactivity or an increased rate of adaptation to stimulation. No direct evidence bearing on this point is offered by the present study. However, the results pertaining to the Anxiety Differential indicate that the psychopath does not evidence much anxiety even after experiencing an intense arousal treatment. If one can assume that there is a positive relationship between physiological arousal and anxiety level immediately following an intensive stressful experience, then it would seem that the psychopath is indeed an autonomically unreactive individual.

³ S. Alexander, personal communications, July 1966.

REFERENCES

- BARRON, F. Complexity-simplicity as a personality dimension. *Journal of Abnormal and Social Psychology*, 1953, 48, 163-172.
- BERLYNE, D. C., & LEWIS, J. L. Effects of heightened arousal on exploratory behavior. *Canadian Journal of Psychology*, 1963, 17, 398-411.
- HARE, R. D. Psychopathy, fear arousal and anticipated pain. *Psychological Reports*, 1965, 16, 499-502.
- HAYWOOD, H. C. Novelty seeking behavior as a function of manifest anxiety and physiological arousal. *Journal of Personality*, 1962, 30, 63-74.
- HUSEK, T., & ALEXANDER, S. The effectiveness of the Anxiety Differential in examination stress situations. *Educational and Psychological Measurement*, 1963, 23, 309-318.
- LYKKEN, D. T. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, 1957, 55, 6-10.
- MCREYNOLDS, P., & BRYAN, J. Tendency to obtain new percepts as a function of the level of un-

- assimilated percepts. *Perceptual and Motor Skills*, 1956, 6, 183-186.
- MUNSINGER, H., & KESSEN, W. Uncertainty, structure, and preference. *Psychological Monographs*, 1964, 78(9, Whole No. 586).
- O'LEARY, K. S. Preference for variability of stimuli as a function of experimentally induced anxiety. *Psychological Reports*, 1965, 16, 1202.
- PETERSON, D. R. Behavior problems of middle childhood. *Journal of Consulting Psychology*, 1961, 25, 205-209.
- PETERSON, D. R., QUAY, H. D., & CAMERON, G. R. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. *Journal of Consulting Psychology*, 1959, 23, 395-399.
- PETERSON, D. R., QUAY, H. C., & TIFFANY, T. C. Personality factors related to juvenile delinquency. *Child Development*, 1961, 32, 355-372.
- QUAY, H. C. Dimensions of personality in delinquent boys as inferred from the factor analysis of behavior ratings. *Journal of Research in Crime and Delinquency*, 1964, 1, 33-37.
- QUAY, H. C. Psychopathic personality as pathological stimulus seeking. *American Journal of Psychiatry*, 1965, 122, 180-183.
- QUAY, H. C., SPRAGUE, R. L., SHULMAN, H. S., & MILLER, A. L. Some correlates of personality disorder in a child guidance clinic sample. *Psychology in the Schools*, 1966, 3, 44-47.
- SIEVEKING, N. A. Effects of age on preference for incongruity. Unpublished master's thesis, University of Illinois, 1965.
- SKRZYPEK, G. J. The effect of perceptual isolation and arousal on anxiety, complexity preference and novelty preference in psychopathic and neurotic delinquents. Unpublished doctoral dissertation, University of Illinois, 1967.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received November 20, 1967)

CONFLICT AND DOMINANCE WITHIN FAMILIES OF DISTURBED CHILDREN¹

JOSEPH BECKER²

AND

EILEEN IWAKAMI

University of Washington

University of Illinois

Twenty-six sets of clinic parents with sons referred for psychological treatment were compared with 42 sets of nonclinic controls in a structured interaction situation from which conflict and dominance indices were derived. Crucial concerns were the internal consistency of the indices, whether they measured unitary constructs, and whether amount of verbal activity appreciably effected the frequency-type conflict measures. Split-half reliabilities of most indices were quite adequate; intercorrelations between dominance indices were considerably lower than for conflict indices. But the latter were strongly related to amount of verbal activity. Interaction indices were subjected to 2×2 (clinic, non-clinic; child present, absent) analyses of covariance. Nonclinic parents exhibited somewhat more conflict than clinic families, but adjustment of scores for verbal activity considerably attenuated these differences.

Though family interaction has long been implicated as a crucial etiological factor in personality disturbance, systematic family investigations are a recent development. Consequently, methodological approaches are in a formative stage. Fontana's (1966) excellent critique of family etiological studies strongly endorses direct observational approaches. A structured interactional variant of the direct approach based on work by Strodbeck (1951) has been used in several recent studies of parental dominance and conflict (e.g., Becker & Siefkes, 1969; Farina, 1960; Farina & Dunham, 1963; Hetherington, 1965). These investigations did not assess whether the indices used to estimate dominance and conflict measure unitary constructs, nor the vulnerability of the frequency conflict indices to amount of verbal activity, nor the internal reliability of the indices.

The present study sought to clarify these measurement issues within the context of determining whether discriminable patterns of conflict and dominance characterize the parents of children referred for psychological treatment from parents of nonclinic controls.

A "startling" dearth of studies on relations between parental interaction and child behavior has been noted by Hoffman and Lippett (1960, pp. 960-961). Clinicians tend to assume that relations between the parents of behaviorally disturbed children are more conflictual and less equalitarian than those between the parents of better adjusted children. Several clinical (e.g., Lidz, Cornelison, Terry, & Fleck, 1958; Wynne, Ryckoff, Day, & Hirsch, 1958) and experimental studies (Farina, 1960; Lerner, 1965) of interaction between the parents of *psychotically* disturbed children support these assumptions. But findings on the parents of less severely disturbed children are less consistent (e.g., Bandura, 1960; Baruch & Wilcox, 1944; Burchinal, Hawkes, & Gardner, 1957; Krug, 1965; McCord, McCord, & Howard, 1961). To what extent disparities between results of these studies are due to such variations as children's ages, social class sampling, and/or methodological approach is indeterminate at present.

Initially, it was planned to subdivide clinic children into groups with predominantly conduct (acting out, aggressive) or personality (shy, withdrawn, anxious) disturbances. But since most clinic children displayed mixed

¹ This study was supported by United States Public Health Service Grant MH-06781, from the National Institute of Mental Health and Grant 41-32-66-306 from the University of Illinois Research Board. Appreciation is expressed to Gary F. Beanblossom and Gerald Edwards of the University of Washington Testing Bureau and to Ethel Shaeffer of the Biostatistics Division of the University of Washington Medical School for assistance in computer analysis of these data. Portions of this paper were presented at the 1967 Western Psychological Association Meeting.

² Requests for reprints should be sent to Joseph Becker, Department of Psychiatry, School of Medicine, University of Washington, Seattle, Washington 98105.

types of behavioral deviance as gauged by clinical evaluation and objective rating scales, families of clinically referred children were pooled in the data analyses.

METHOD

Subjects

Clinic subjects (*Ss*) were males referred to the Psychological Clinic of the University of Illinois and their parents. Twenty-six consecutive admissions who met the *S* criteria were used. Child criteria included age 5-12, Caucasian, and no mental retardation, organic brain dysfunction, psychosis, or serious physical impairment. Both biological parents had to be available, and able to meet intellectual, mental status, and organic requirements similar to those specified for their children.

Nonclinic controls were solicited by mail and paid \$10.00. Prospective volunteers were selected randomly from the rolls of schools attended by the clinic sample. Slightly over 20% of the families written to participated. Nonclinic child-parent groups met the same criteria as the clinic patient-parent groups except that families with histories of serious behavioral disturbance were eliminated (five).

The surnames of both groups appeared to reflect proportionately the predominantly British and German backgrounds of the regional population.

Procedure

The three family members (mother, father, son) were seen simultaneously, but *individually*, during the *first phase* of the experimental session.

Each parent was seen briefly for socioeconomic data (Hollingshead, 1957), critical social history, and administration of the Wechsler Adult Intelligence Scale (WAIS), Vocabulary subtest. Parents then privately completed the following: (a) Behavior Rating Schedule (BRS) (Becker & Krug, 1964), ratings were obtained on the male child *S*, his sibs, self, and spouse; (b) the Femininity subscale of the California Psychological Inventory (CPI); (c) the Franck Drawing Completion Test (FDCT); (d) and a Child Problem Questionnaire (CPQ) (Jackson, 1956), consisting of 12 hypothetical problem situations. These CPQ items require an indication of how the parent would cope with his child's aggression. Finally, a 5-min. spontaneous speech sample (SSS) about family life was tape-recorded.

During the *individual phase* of the child's testing, he first indicated in counterbalanced order how one of his parents would react to the CPQ items. Then the It Test of sex-typing, a five card TAT measure of covert sex identification, and the Vocabulary and Block Design subtests of the Wechsler Intelligence Scale for Children (WISC) were given. Next, he indicated how his other parent would react to the

CPQ problems and concluded by recording a 5-min. speech sample.³

In the subsequent *joint phase*, parents and child were brought together. Both parents were present throughout the joint session; the child for only one-half. Family members were again presented with the CPQ. Problems were read one at a time by the experimenter (*E*). Each situation was discussed until a solution was reached. A counterbalanced design determined which half of the session the child would be present in, and whether odd or even numbered problems would be presented first. Joint sessions were tape-recorded.

To the extent feasible, the *Es* did not know whether families were clinic or nonclinic ones. Data were coded so that scorers did not know their source.

Interactional Indices

The indices of conflict and dominance derived from the joint discussion of the CPQ were taken from Farina (1960).⁴ Farina (1960) reported the interrater reliabilities to be generally satisfactory for the less reliable indices (ranging from .74 to .96; overall average rho reliability = .89). Interrater reliabilities on five cases for the same indices in the present study were slightly less (range .70 to .97; overall rho = .84). In addition, odd-even reliability coefficients for the conflict-dominance indices were obtained. These Spearman-Brown corrected reliability coefficients are given after each index below. All *r*'s were significant ($p < .01$, $N = 68$) except for Yielding-Maximum. A test-retest study of related family interaction variables by Ferreira, Winters, and Poindexter (1966) using a 6-mo. interval yielded fairly satisfactory correlations ranging from .57 to .76.

Indices of conflict used were (a) Frequency of Simultaneous Speech (.81) (Freq. Simult. Sp.); (b) Duration of Simultaneous Speech (.84) (Durat. Simult. Sp.); (c) Interruptions by Father (.63) (Interrupt.-Fa.); (d) Interruptions by Mother (.57) (Interrupt.-Mo.); (e) Interruptions-Total (Interrupt.-Tot.); (f) Disagreements and Aggressions—Father (.65) (Disagr. & Aggr.-Fa.); (g) Disagreements and Aggressions—Mother (.56) (Disagr. & Aggr.-Mo.); (h) Disagreements-Total (Disagr.-Tot.); (i) Failure to Agree (.60) (Fail. Agree); (j) Verbal Activity (.87) Verb Act..

Indices of dominance were (a) Speaks first (.67); (b) Speaks Last (.79); (c) Speaks First and Last-Total (Spk. 1st & L-Tot.); (d) Passive Acceptance of Solution (.65) (Pass. Acc. Sol.); (e) Total Time Spoken (.87) (Tot. T. Spkn.); (f) Yield—Maximum (.15) (Yld.-Max.); (g) Yield—Minimum (.40) (Yld.-Min.).

³ The results for BRS sib, self, and spouse ratings, CPI, FDCT, SSS, and It are not discussed because these measures are not directly relevant to the focus of this paper.

⁴ A more detailed discussion of the rationale and scoring of indices is available in Farina (1960).

RESULTS

Supplementary Variables

Clinic and nonclinic groups were compared on the concomitant variables of age, estimated IQ, education, and socioeconomic class. Estimated IQs and interdependent educational and social class levels of the nonclinic parents were higher than those of the clinic parents.⁵ The mean social class score for both groups fell within Hollingshead's Level III, but at opposite ends of the score range.

Intercorrelations between Concomitant and Dependent Variables

Since the pattern of intercorrelations for the clinic and nonclinic parents did not differ essentially, the groups were pooled with the results shown in Table 1. The parental supplementary variables on which the groups differed, IQ, education, and socioeconomic level, were highly intercorrelated. These three supplementary variables were positively related to the conflict variables: Freq. Simult. Sp., Durat. of Sp., and Verb. Act. for both fathers and mothers, and to Interrupt. for mothers. The supplementary variables were unrelated to the dominance indices except that brighter mothers showed less passive acceptance.

Intercorrelations between Dependent Variables

For both fathers and mothers, the conflict variables were positively related to each other with the exception of Fail. Agree. For fathers, conflict variables were unrelated to dominance variables except for Verb. Act. which correlated positively with three of the four dominance variables. For mothers also, conflict and dominance variables were unrelated except that Pass. Accept. Sol. was negatively related to conflict measures. Unlike the conflict variables, the dominance variables were unrelated to each other except for Pass. Accept. Sol. which was negatively correlated with Speaks First and Last. The extent of the intercorrelations between Verbal Activity and

the other conflict measures should be noted especially.

The intercorrelations of Table 1 yielded stronger internal consistency support for a dimension of conflict than of dominance. Because of the high intercorrelation between the supplementary variables and the similarity of their relations to the dependent variables, scores in subsequent analyses were adjusted for only the supplementary variable of socioeconomic class.

Analyses of Conflict Indices

The scores for each conflict index were initially analyzed by a 2×2 (clinic, non-clinic; child present, child absent) repeated measures analysis of covariance adjusting for socioeconomic class and using an unweighted-means solution for the unequal group sizes (Winer, 1962, pp. 374-378). Several trends were apparent. Nonclinic parents consistently scored higher on the conflict variables and significantly so on Freq. Simult. Sp. ($p < .01$), Interrupt.-Fa. ($p < .05$), Interrupt.-Total ($p < .05$), and Verb. Act. ($p < .005$). The present-absent effect was significant only for Disagr. & Aggr.-Fa. ($p < .025$), Disagr. & Aggr.-Tot. ($p < .025$), and Verb. Act. ($p < .001$), with conflict always higher under absent conditions. The only significant interaction was on Verb. Act. ($p < .05$). This interaction resulted from a sizable increase in verbal activity during the child's absence by nonclinic parents, whereas the verbal output of clinic parents remained relatively constant. Nonclinic parents tended to be more conflictual in the child's absence on every conflict variable but one (Fail. Agree), whereas no pattern was evident for the clinic parents.

Interpretation of the results on these frequency measures is complicated by the role of verbal activity. Multiple covariate analyses of variance adjusting the log transformed conflict scores for both socioeconomic class and verbal activity resulted in a significant clinic-nonclinic difference on Freq. Simult. Sp. ($p < .05$) only. Absence of the child was related to higher conflict between the parents on Disagr. & Aggr.-Fa. ($F = 9.11$, $df = 1/65$, $p < .005$), Disagr. & Aggr.-Mo. ($F = 4.73$, $df = 1/65$, $p < .05$), and Disagr. & Aggr.-

⁵ An extended report is available from the senior author containing tables of these and other data from this study.

TABLE 1
INTERCORRELATIONS BETWEEN CONCOMITANT VARIABLES AND INDICES OF CONFLICT AND DOMINANCE FOR
COMBINED CLINIC AND NONCLINIC FATHERS AND FOR MOTHERS

Variables	Fathers (N = 68)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age		.16	.13	-.14	-.05	-.03	-.11	-.17	.23	-.13	-.02	-.07	-.10	.17
2. Est. IQ	.24		.80**	.81**	.30*	.31*	.14	.15	-.06	.41**	.18	.10	-.19	-.09
3. Education	.12	.66**		-.93**	.33**	.32*	.21	.19	-.10	.29*	.17	.03	.00	.01
4. Socio-Ec. Class ^a	-.13	-.56**	-.70**		-.37**	-.37**	-.23	-.16	.11	-.34**	-.16	-.06	.09	.00
5. Freq. Simult. Sp.-C ^b	-.12	.19	.33**	-.37**		.93**	.75**	.50**	-.08	.42**	.01	.01	-.24	.00
6. Durat. Simult. Sp.-C	-.04	.26*	.31*	-.37**	.93**		.65**	.36**	-.09	.37**	.00	.05	-.18	-.02
7. Interrupt. -C	-.01	.25*	.37**	-.39**	.82**	.74**		.68**	-.06	.38**	.06	-.08	-.23	-.09
8. Disagr. & Aggr. -C	-.17	-.05	.12	-.07	.49**	.39**	.54**		.10	.30*	-.03	.07	-.18	-.14
9. Fail. Agree -C	.17	.12	.07	.11	-.08	-.09	-.08	.20		.12	-.32*	-.18	-.11	-.22
10. Verb. Act. -C	-.13	.37**	.33**	.34**	.47**	.40**	.43**	.41**	.08		.31*	.35**	-.33**	-.05
11. Speaks First -D ^c	-.02	.01	-.08	.17	-.01	.00	-.05	.03	.33**	-.02		.22	-.27*	.20
12. Speaks Last -D	.04	.17	-.01	-.06	-.01	-.05	.12	.07	.18	.19	.24		-.32*	.17
13. Pass. Accept. Sol.-D	.13	-.33**	-.15	.11	-.38**	-.36**	-.41**	-.27*	-.16	-.47**	-.25*	-.28*		-.16
14. Tot. T. Spkn. -D	.15	-.01	.11	.00	.00	-.02	-.10	-.17	-.22	-.05	-.21	-.16	-.03	

Note.—Correlations above the diagonal are for fathers, those below are for mothers.

^a High scores are inversely related to social class.

^b C signifies a conflict variable.

^c D signifies a dominance variable.

* $p < .05$.

** $p < .01$, two-tailed tests.

Tot. ($F = 8.52$, $df\ 1/65$, $p < .005$). None of the interactions approached significance.

Analyses of Dominance Indices

Scores on the dominance indices were also analyzed by a 2×2 repeated measures analyses of covariance (Winer, 1962, p. 312) with scores adjusted for group differences in socioeconomic class. None of the F tests for main effects or interaction on any of the dominance variables were significant. Fathers tended to be dominant on all dominance variables for both groups with nonclinic fathers consistently, though not significantly, more dominant than clinic fathers.

Behavior Rating Schedule

Parents' ratings on each of the subscales of the BRS were subjected to an analysis of covariance adjusting the scores of each group for the parents' estimated IQ and socioeconomic class. Clinic mothers (CM) and fathers (CF) rated their children's behavior as significantly more deviant than the control mothers (NCM) and fathers (NCF). Parental ratings of children in both groups were highly correlated (r 's for all subscales significant at $p < .01$ for all subscales in both groups).

Clinic mothers rated themselves as significantly more deviant than nonclinic mothers, whereas clinic and nonclinic fathers' self-ratings did not differ. Clinic mothers rated themselves as more deviant than their husbands, whereas the nonclinic parents' self-ratings did not differ.

DISCUSSION

Despite their face validity, the intercorrelations among the dominance indices afford negligible support for viewing them as measures of a unitary construct. While partialing out the effects of verbal activity on intercorrelations between the frequency-type conflict indices attenuates their relatedness, moderate support does remain for their tapping a common construct. Since most of the dominance and conflict indices have respectable split-half and interrater reliability, further investigation of their validity and correlates is indicated. Previous studies using these

indices without correction for the effects of verbal activity must be interpreted cautiously. If Farina's (1960) finding of greater verbal activity when parents cannot agree is replicated, it would be useful to evaluate conflict findings with and without verbal activity partialled out. The present study had too few failures to agree to meaningfully assess this problem.

Evaluation of the substantive findings of this study is hampered by the possibilities of sampling bias inherent in the 20% volunteer rate for controls.

The finding that nonclinic parents interacted somewhat more conflictually supports the contention that qualitative aspects of intrafamilial conflict may be more related to psychopathology than quantitative ones (Spiegel, 1957; Wynne et al., 1958). These investigators argue that realistic adaptation is fostered by exposure to the explicit identification of distinctive viewpoints, their articulation, mutual respect for differences and constructive conflict resolution. Empirical support for these theoretical speculations are indirectly provided by several studies (Ferreira, Winter, & Poindexter, 1966; Fisher, Boyd, Walker, & Sheer, 1959; Singer & Wynne, 1963). As with anxiety, it appears unlikely that amount of conflict per se bears any simplistic functional relation to severity of psychopathology. Why greater conflict occurred between the nonclinic parents in the child's absence is unclear. Possibly nonclinic parents exercise greater judgment and considerateness of their children by reducing their verbal activity which permits the child greater expression.

In conclusion, the BRS received further validation as a classification instrument, and previous MMPI findings of more self-reported disturbance in clinic mothers than fathers or controls (Goodstein & Rowley, 1961) were supported. The relatively high clinic interparental correlations on child ratings did not support etiologic notions of neuroses as based on discrepant parental perceptions of the child.

REFERENCES

- BANDURA, A. *Relationship of family patterns to child behavior disorders.* (NIMH Progress Report M-

- 1734) Washington, D. C.: United States Government Printing Office, 1960.
- BARUCH, D. W., & WILCOX, J. A. A study of sex differences in preschool children's adjustment co-existent with interparental tensions. *Journal of Genetic Psychology*, 1944, **64**, 281-303.
- BECKER, J., & SIEPKES, H. Parental dominance, conflict, and disciplinary coerciveness in families of female schizophrenics. *Journal of Abnormal Psychology*, 1969, **74**, 193-198.
- BECKER, W. C., & KRUG, R. S. A circumplex model for social behavior in children. *Child Development*, 1964, **35**, 371-396.
- BURCHINAL, L. G., HAWKES, G. R., & GARDNER, G. Marriage adjustment, personality characteristics of parents and the personality adjustment of their children. *Marriage and Family Living*, 1957, **19**, 366-372.
- FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology*, 1960, **61**, 31-38.
- FARINA, A., & DUNHAM, R. M. Measurement of family relationships and their effects. *Archives of General Psychiatry*, 1963, **9**, 64-73.
- FERREIRA, A. J., WINTER, W. D., & POINDEXTER, E. J. Some interactional variables in normal and abnormal families. *Family Process*, 1966, **5**, 60-75.
- FISHER, S., BOYD, I., WALKER, D., & SHEER, D. Parents of schizophrenics, neurotics, and normals. *Archives of General Psychiatry*, 1959, **1**, 149-167.
- FONTANA, A. F. Family etiology of schizophrenia: Is a scientific methodology possible. *Psychological Bulletin*, 1966, **66**, 214-228.
- GOODSTEIN, L. D., & ROWLEY, V. N. A further study of parents of disturbed and nondisturbed children. *Journal of Consulting Psychology*, 1961, **25**, 460.
- HETHERINGTON, E. M. A developmental study of the effects of sex of the dominant parent on sex-role preference, identification, and imitation in children. *Journal of Personality and Social Psychology*, 1965, **2**, 188-195.
- HOFFMAN, L. W., & LIPPETT, R. The measurement of family life variables. In P. H. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960.
- HOLLINGSHEAD, A. B. *Two factor index of social position*. New Haven: Yale University, 1957. Available from the author.
- JACKSON, P. W. Verbal solutions to parent-child problems. *Child Development*, 1956, **27**, 339-349.
- KRUG, R. S. Attitudinal and behavioral differences in childrearing practices between parents of clinic and non-clinic children: A questionnaire assessment. Unpublished doctoral dissertation, University of Illinois, 1965.
- LENER, P. M. Resolution of intrafamilial role conflict in families of schizophrenic patients. I: Thought disturbance. *Journal of Nervous and Mental Disease*, 1965, **141**, 342-351.
- LIDZ, T., CORNELISON, A., TERRY, D., & FLECK, S. Intrafamilial environment of the schizophrenic patient. VI: The transmission of irrationality. *Archives of Neurology and Psychiatry*, 1958, **79**, 305-316.
- MCCORD, W., MCCORD, J., & HOWARD, A. Familial correlates of aggression in nondelinquent male children. *Journal of Abnormal and Social Psychology*, 1961, **62**, 79-93.
- SINGER, M. T., & WYNNE, L. C. Differentiating characteristics of parents of childhood schizophrenics, childhood neurotics, and young adult schizophrenics. *American Journal of Psychiatry*, 1963, **120**, 234-243.
- SPIEGEL, J. P. The resolution of role conflict within the family. *Psychiatry*, 1957, **20**, 1-16.
- STRODTBECK, F. L. Husband-wife interaction over revealed difference. *American Sociological Review*, 1951, **16**, 468-473.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WYNNE, L. C., RYCKOFF, I. M., DAY, J., & HIRSCH, S. I. Pseudomutuality in the family relations of schizophrenia. *Psychiatry*, 1958, **21**, 205-220.

(Received December 9, 1967)

EYELID CONDITIONING AND PERSONALITY: POSITIVE RESULTS FROM NONPARTISANS

ELLEN V. PIERS¹ AND ELIZABETH P. KIRCHNER²

Pennsylvania State University

The Ss selected on the basis of either high- or low-response probability on the last 60 of 300 conditioning trials obtained significantly different scores on measures of manifest anxiety, extraversion, and neuroticism, and significant correlations occurred among these personality measures. The results are discussed in terms of the theoretical positions of Spence and Eysenck.

Investigators concerned with relationships between eyelid conditioning and personality variables have typically offered support for the hypotheses of either Spence (e.g., Taylor, 1951) or Eysenck (e.g., Franks, 1957), or they have reported negative findings (e.g., Franks, 1963; King, Kimble, Gorman, & King, 1961) which the appropriate theorist has claimed resulted from the absence of certain necessary conditions.

K. W. Spence alone (1964) and in collaboration (J. T. Spence & K. W. Spence, 1966; K. W. Spence & J. T. Spence, 1964, 1966), working within the Hullian framework, has argued for a positive association between conditionability and anxiety, as measured by the Manifest Anxiety (*MA*) scale (Taylor, 1951). This relation is derived from the assumption that the degree to which an individual admits to characteristically exhibiting manifest symptoms of anxiety, as described by items on the scale, will be related to the magnitude of his emotional responsiveness and therefore to the level of drive (*D*) in a conditioning situation. K. W. Spence considered that results from numerous studies at Iowa and elsewhere generally supported his position. He explained the failure of other investigators to reproduce his positive results as a function of such variables as the following: too small a sample, nonnaïve Ss, laboratory conditions which failed to arouse emotionality, and inclusion of voluntary responders. He and J. T. Spence (1964) re-

ported high correlations between the *MA* scale and Eysenck's Neuroticism (*N*) scale and positive correlations between *N* and conditioning but no significant correlations between Eysenck's Extraversion (*E*) scale and conditioning.

Eysenck (1965) has hypothesized that conditioning is correlated positively with introversion (or negatively with extraversion) on the basis of Pavlovian theory that cortical inhibition depresses conditioning, and from his assumption that cortical inhibition is characteristic of extraverts, while cortical excitation is characteristic of introverts. Using his Maudsley Personality Inventory (1962) which gives *E* and *N* scores, he and his colleagues reported significant negative correlations between conditioning and *E* which he felt were a function of the fact that experimental conditions conducive to inhibition have been met. These conditions included (*a*) partial as compared with complete reinforcement, (*b*) weak as opposed to strong CS and UCS, and (*c*) discrimination learning as opposed to single stimulus conditioning. He reported little or no correlation between conditioning and *N*.

Eysenck has acknowledged that *MA* scale scores may, on occasion, correlate positively with conditioning performance. He previously (1957) hypothesized that such a relationship comes about because of the correlation of the *MA* scale with extraversion rather than because of the observed correlation of the *MA* scale with neuroticism. This position followed from his theory that neuroticism and extraversion are orthogonal dimensions of personality, with conditioning related to the extraversion-introversion dimension but not to neuroticism. More recently (1965) he has tried

¹ Requests for reprints should be made to Ellen V. Piers, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

² The authors wish to thank David Higgins for his kindness in making available the conditioning data.

to incorporate Spence's findings by distinguishing between autonomic arousal (brought about by emotion-producing conditions) as opposed to cortical activation (excitation) which is related to the E-I dimension. According to this position, he predicted that autonomic arousal results in correlations between N and conditioning while inhibition-producing conditions lead to negative correlations between conditioning and E and little or no relation between conditioning and N.

The purpose of the present study was to throw further light on these two positions from a neutral theoretical viewpoint, taking into account the specific experimental conditions proposed by both theorists. In contrast to most studies which have investigated personality-eyelid conditioning variables by selecting Ss on the basis of their personality scores and reporting the subsequent performance of these Ss on 50 or fewer conditioning trials, the present study utilized Ss selected on the basis of their eyelid conditioning over 300 trials and reports subsequent performance on personality measures.

METHOD

Subjects

The original Ss were 270 students of both sexes who had participated in an eyelid-conditioning study while enrolled in the introductory psychology course at Pennsylvania State University. No requirements were made as to naiveté concerning psychology experiments. The Ss for the present study were 45 of the highest conditioners and 34 of the lowest conditioners from the original pool. Fifty-five of the Ss were male and 24 were female. This is approximately the ratio of males to females in the University undergraduate population.

The Conditioning Study

The conditioning apparatus and procedure are described elsewhere (Prokasy, Carlton, & Higgins, 1967). For purposes of the present study it is sufficient to report that (a) the CS and UCS can be considered relatively weak, (b) CS-UCS intervals of 360, 500, and 860 msec. were used and found not

to be significantly related to mean response probability; hence data from the three CS-UCS intervals were combined, (c) there was a 50% partial reinforcement schedule subdivided into three patterns of reinforcement which are irrelevant to the present study except insofar as it was necessary to assure equal representation of each pattern, (d) there was no attempt to arouse emotion, either by the apparatus, by surroundings, or by the taped instructions which instructed Ss simply to sit quietly and let their reactions take care of themselves, and (e) for each S the 300 conditioning trials took the better part of 1 hr.

Procedure

For the present study the 270 Ss were ranked on the basis of their mean response probability on the last 60 of the 300 conditioning trials, and attempts were made to find current addresses for Ss in roughly the upper and lower quartiles of this distribution. This method resulted in the formation of two groups, each with 63 potential Ss, who were sent the MA scale and the Maudsley Personality Inventory (MPI) with a cover letter explaining the project and offering them \$1.25 for completion of the scales. Forty-five high conditioners (HC) and 34 low conditioners (LC) returned the completed scales. (Inability to assume that the material actually reached all potential Ss precludes any attempt to interpret the differential rate of return.) For the HC group the mean response probability of the last 60 trials was $.89 \pm .09$; for the LC group the probability was $.18 \pm .14$.

RESULTS

Before analyzing the scores of the HC and LC Ss, means and standard deviations on the personality variables for the total group and for each sex were computed and found to be similar to other United States college student populations (Knapp, 1962; Spence, Farber, & Taylor, 1954). These data are shown in Table 1. Since mean sex differences were found to be nonsignificant on all measures, data for both sexes combined were used in subsequent analyses.

Table 2 presents the MA scale and MPI means of the HC and LC Ss. It can be seen that HCs were characterized by significantly higher scores on all three personality vari-

TABLE 1
MEANS AND STANDARD DEVIATIONS OF Ss ON PERSONALITY MEASURES

Measure	Males (<i>N</i> = 55)	Females (<i>N</i> = 24)	Total (<i>N</i> = 79)
Manifest Anxiety scale (MA scale)	13.18 ± 7.58	14.96 ± 8.76	13.72 ± 7.94
Extraversion (MPI)	30.29 ± 9.37	28.96 ± 10.62	29.89 ± 9.72
Neuroticism (MPI)	20.51 ± 10.39	21.04 ± 10.95	20.67 ± 10.49

TABLE 2
COMPARISONS OF PERSONALITY SCORES FOR PRE-SELECTED HIGH AND LOW CONDITIONERS

Variable	Group		<i>t</i>
	HC (<i>N</i> = 45)	LC (<i>N</i> = 34)	
Manifest Anxiety scale	16.04 ± 8.53	10.65 ± 5.91	3.32**
MA scale	28.02 ± 10.84	32.35 ± 7.45	2.10*
Extraversion (MPI)	24.18 ± 10.54	16.03 ± 8.55	3.79**
Neuroticism (MPI)			

* *p* < .05 (one- or two-tailed test).
** *p* < .01 (one- or two-tailed test).

ables, although differences between E scores were significant at a lower level (*p* < .01 for MA scale and N, *p* < .05 for E).

Table 3 presents the intercorrelations between conditioning and the personality variables. Inspection shows that all correlations were significant beyond the .01 confidence level except that for Conditioning and Extraversion, which was significant only beyond the .05 level. It must be remembered that these correlations are based on extreme scores which were used in order to make the results comparable to many of the Spence studies. It is possible, therefore, that correlations over

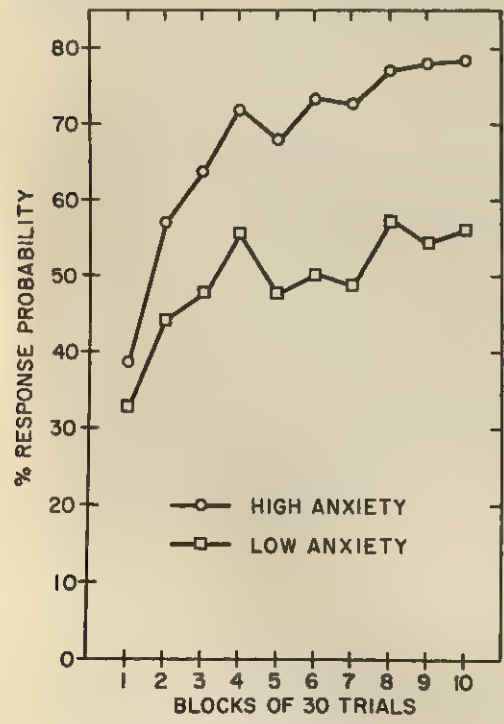


FIG. 1. Conditioning scores for high- and low-anxiety Ss.

the whole distribution of conditioning scores might be slightly lower.

To determine the degree to which the correlation between the MA scale and conditioning might be attributed to the relation between these variables and E, a partial *r* was computed, removing the effects of E. The original *r* of .30 between conditioning and MA scale was thereby reduced to .24 which was still significant at the .05 level. On the other hand, when a similar manipulation removed the effects of N, the correlation between MA scale and conditioning dropped to .06, which was not significantly different from zero.

One of the unique features of the present study was the very large number of conditioning trials. To test statements (e.g., J. T. Spence & K. W. Spence, 1966) that frequency curves for high- and low-anxiety levels will gradually diverge over trials and also to observe the effect of a longer experimental period on the two levels, conditioning scores were plotted for Ss in the upper and lower 20% of the MA scale distribution (raw scores 21 or over and 7 or under, respectively). These curves are shown in Figure 1. It can be seen that divergence did in fact occur in the course of the early trials and that these differences then remained fairly constant throughout the whole conditioning period, while the mean response probabilities appeared to reach asymptote late in the period, after approximately 240 trials. (The same general effect, but less marked, was evident when conditioning scores of Ss in the upper and lower 20% of the E distribution were plotted.)

DISCUSSION

The standard conditioning procedures to which Ss of this study were exposed were not

TABLE 3
CORRELATIONS AMONG CONDITIONING AND PERSONALITY MEASURES

Measure	Conditioning	MA scale	E
Manifest Anxiety Scale			
MA scale	.305**		
Extraversion (MPI)	-.223*	-.399**	
Neuroticism (MPI)	.341**	.794**	-.390**

* *p* < .05.
** *p* < .01.

designed to be either emotion arousing or inhibition producing. The large original pool of Ss did permit comparisons of an adequate sample at the extremes of conditionability. Also, relatively weak CS and UCS were employed under partial reinforcement and massed trials. Other variables specified by Spence or Eysenck were not considered in the selection and treatment of Ss.

Under this empirical approach, results supported the relationship between conditioning and both Extraversion and Anxiety (Neuroticism). In other words, without attempting to fulfill most of Spence's conditions, a correlation was obtained higher than he, himself, has reported (1964) for conditioning and the MA scale. Contrary to his findings, a significant correlation between conditioning and E was also found.

Eysenck (1965) has suggested that under ordinary conditions of eyelid conditioning, correlations with neither E nor N are likely to appear. Under inhibition producing conditions he reported no significant correlations between conditioning and N. Since positive correlations were found for both E and N, the present findings do not support either of these positions. Further, the substantial negative relation between E and N lends weight to the accumulating evidence (Jensen, 1965) that for college students, as well as psychiatric populations, these dimensions can hardly be considered orthogonal. Finally, the correlation between MA scale and conditioning does not appear to be due, as Eysenck hypothesized earlier, to their relationship with Extraversion.

Although Spence's main position is supported to a much greater degree than is Eysenck's, it would seem that neither Eysenck's nor Spence's recent explanations can account for all the significant relationships found in this study. These positive data present a different problem than do negative data in testing a theory. The failure to find relationships where they are predicted can be ignored or explained away on the grounds that necessary conditions have not been met. On the other hand, when significant relationships are obtained empirically and persist over some

period of time, in the absence of some of the proposed conditions, they must be acknowledged. Existing theories would thus appear to need modification to incorporate findings such as are presented here.

REFERENCES

- EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.
- EYSENCK, H. J. *Maudsley Personality Inventory*. San Diego, Calif.: Educational & Industrial Testing Service, 1962.
- EYSENCK, H. J. Extraversion and the acquisition of eyeblink and GSR conditioned responses. *Psychological Bulletin*, 1965, **63**, 258-270.
- FRANKS, C. M. Personality and eyeblink conditioning seven years later. *Acta Psychologica*, 1963, **21**, 295-312.
- FRANKS, C. M. Personality factors and the rate of conditioning. *British Journal of Psychology*, 1957, **48**, 119-126.
- KING, M. S., KIMBLE, G. A., GORMAN, J., & KING, R. A. Replication report: Two failures to reproduce effects of anxiety on eyelid conditioning. *Journal of Experimental Psychology*, 1961, **62**, 532-533.
- KNAPP, R. R. *Manual for the Maudsley Personality Inventory*. San Diego, Calif.: Educational & Industrial Testing Service, 1962.
- JENSEN, A. R. Review of Maudsley Personality Inventory. In O. K. Buros (Ed.), *The sixth mental measurements yearbook*. Highland Park, N. J.: Gryphon Press, 1965.
- PROKASY, W. F., CARLTON, R. A., & HIGGINS, J. D. Effects of non-random intermittent reinforcement schedules in human-eyelid conditioning. *Journal of Experimental Psychology*, 1967, **74**, 282-288.
- SPENCE, J. T., & SPENCE, K. W. The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press, 1966.
- SPENCE, K. W. Anxiety (drive) level and performance in eyelid conditioning. *Psychological Bulletin*, 1964, **61**, 129-139.
- SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *Journal of Experimental Psychology*, 1954, **48**, 404-408.
- SPENCE, K. W., & SPENCE, J. T. Relation of eyelid conditioning to manifest anxiety, extraversion, and rigidity. *Journal of Abnormal and Social Psychology*, 1964, **68**, 144-149.
- SPENCE, K. W., & SPENCE, J. T. Sex and anxiety differences in eyelid conditioning. *Psychological Bulletin*, 1966, **65**, 137-142.
- TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *Journal of Experimental Psychology*, 1951, **41**, 81-92.

(Received December 15, 1967)

REMEMBRANCE OF FEELINGS PAST: A STUDY OF PHENOMENOLOGICAL GENETICS¹

ANNE-MARIE ALLERAND BLOCH²

Teachers College, Columbia University

The hypothesis that subjective experience may be partially genetically determined was tested with a sample of 24 pairs of monozygotic (MZ) twins, 24 pairs of dizygotic (DZ) twins, and 24 pairs of singleton siblings (Sibs). The verbal reports of their experiences of eight emotional states were content analyzed, and the intrapair similarity of the descriptions was evaluated. Since the MZs were significantly more alike than either the DZs or Sibs, who did not differ from each other, the presence of a genetic factor was inferred. The general language patterns of twins were also investigated, with results indicating no greater similarity among MZs than among DZs.

This study is concerned with exploring the hypothesis that the language of emotional experience is partially determined by genetic factors and to this end has compared the reports of emotional experiences written by monozygotic (MZ) twins, dizygotic (DZ) twins, and singleton siblings (Sibs). The role of twinship per se as a possible factor contributing to experiential similarity was also investigated.

A basic assumption is that verbal reports reflect physiological reactions which are presumably characteristic of emotional states. While there is not necessarily a one-to-one relationship between physiological events and verbal descriptions, and while there are obviously many factors likely to interfere with a report of physiological experiences, it nevertheless seems reasonable to assume that, for example, two persons reporting that they feel flushed are experiencing similar physical sensations. The few studies dealing with verbally reported correlates of physiological reactions relevant to emotional states tend to support this basic assumption (Malmo & Shagass,

1949; Mordkoff, 1964; Ruesch & Finesinger, 1943; Theron, 1948).

Although situational (Hunt, Cole, & Reis, 1958) and cognitive (Schachter, 1964) factors have been shown to be determinants of emotional states, it is clear that emotional reactions also involve certain physiological events. Efforts have been made to identify specific physiological responses associated with particular emotional states, but these have not met with consistent success (Arnold, 1945; Ax, 1953; Graham, 1955; Schachter, 1957). Other studies, however, have unequivocally demonstrated individual differences and consistencies in physiological reactions (Duffy, 1946; Lacey & Van Lehn, 1952; Schachter, 1957). Research with both animal and human Ss supports the hypothesis of a genetic basis for these individualistic patterns (Duffy, 1962; Richmond & Lustman, 1955), and there is evidence that certain constitutional factors play a role in determining subjective responses to distress (Schaffer, 1966).

Studies of twins and their families have also found evidence of a genetic basis for certain physiological functions (Duffy, 1962; Jost & Sontag, 1944) as well as for some psychological traits. Of particular relevance is the finding that both sociability (Scarr, 1965) and traits associated with the dimension of introversion-extraversion (Eysenck, 1956; Shields, 1962) have strong genetic determinants. It seems plausible that differences along these dimensions would be reflected in subjective experiences. The demonstration of a neurophysiological basis for introversion-

¹ This study was based on a doctoral dissertation submitted to the Department of Psychology, Teachers College, Columbia University in partial fulfillment of the requirements for the doctoral degree. The author wishes to express her appreciation to the members of her dissertation committee, Rosedith Sitgreaves and David Ricks, and especially to her chairman, Joel R. Davitz, for his encouragement and guidance throughout this research.

² Requests for reprints should be directed to Anne-Marie Allerand Bloch, 5369 Broadway, Oakland, California 94618.

extraversion by Shagass and Kerenyi (1958) is also of considerable interest.

Although it is assumed that an individual's genetic inheritance influences his physiological reactions during emotional states and thus his verbal descriptions of these states, it is not assumed that genes operate independently of environmental influences; they are always interacting with both the genetic and extra-organismic environments. Therefore, it is hypothesized that persons with similar genetic backgrounds may not only tend to be predisposed to certain internal responses, but may also tend to elicit similar environmental reactions. For example, a smiling, active infant is more likely to get attention from adults than a quieter, less responsive one. Since his innate responsiveness (Freedman, 1965) and activity (Scarr, 1966) are constantly being reinforced, perhaps this child comes to value interpersonal relations more and tends to construct his world more in terms of the degree of comfort of these relationships than a less responsive child whose subjective experiences might reflect his lesser interest in people, stemming from his relative lack of experience.

Moreover, assuming that greater similarity of environment leads to more similar experiences and thereby increases the likelihood of similar reactions, twinship per se may be a significant determinant of emotional similarity. For this reason a comparison of the reports of the DZ twins and Sibs was planned since the former are no more alike genetically than the latter, sharing on the average about 50% of their genetic material.

METHOD

Verbal descriptions of eight emotions, Anger, Affection, Delight, Disgust, Excitement, Fear, Sadness, and Worry, were obtained from a sample of MZ twins, DZ twins, and Sibs. These descriptions were content analyzed, and on the basis of this analysis, the intrapair similarity of the descriptions was evaluated. A detailed explanation of the procedure and the system of categorization may be found elsewhere (Allerand, 1967).

Subjects

The Ss, 13-18 yr. of age, came from the San Francisco Bay Area. About 25% of them were identified through one large public school system and a few pairs through Scout leaders. Most, however, were known to someone already participating in the

TABLE 1
AGE (IN YEARS) AND VOCABULARY SCORES

Variable	MZ		DZ		Sib	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	16.0	1.3	15.8	1.3	15.9	1.2
Vocab.*	10.0	2.4	10.4	2.4	10.4	2.5

Note.—*N* = 144 Ss.

**N* = 20 multiple-choice items.

study. All the pairs were of the same sex, and when 12 pairs of males and 12 pairs of females within each of the three groups had agreed to participate, the group was considered closed. The total sample included 24 pairs of MZ twins, 24 pairs of DZ twins, and 24 pairs of siblings, 18 mo. or less apart in age. The Ss, all native-born, white Americans, were on the average 16 yr. old and attended either public or parochial schools. With respect to both age and verbal intelligence the groups were remarkably similar (Table 1). The vocabulary scores fell very close to the norms obtained by Miner (1957), indicating that the present sample was intellectually representative of an average group of high school students. An estimate of their socioeconomic status was made on the basis of parental occupation, when it could be specified. According to a 7-point scale (Warner, Meeker, & Eells, 1949), MZ twins were more heterogeneous than either DZs or Sibs, who were restricted to the top four social classes.

Zygosity Determination

The zygosity of the twins was determined by administering a questionnaire devised by Nichols and Bilbro (1966), which enabled them to determine zygosity with about 93% accuracy when cross-validated against extensive blood typings. While all the twins were actually diagnosed MZ or DZ on the basis of this questionnaire, two genetically determined characteristics were noted at the time of interviewing, earlobe type (free or attached) and the presence or absence of middigit hair. Any differences on these characteristics was to be weighed toward a diagnosis of dizygosity in the event that the questionnaire answers were inconclusive. This did not prove necessary, but it was noted that in no event did discordance on these characteristics occur on questionnaire-diagnosed MZ twins. The four pairs of twins diagnosed monozygotic (identical) according to their hospital records were also found to be identical by the questionnaire.

Instrument and Procedure

The questionnaire consisted of two parts. The first part contained the identifying information (put aside upon completion until the analysis of the data was finished), the zygosity-determination questions, and the material dealing with the subjective experiences. After the zygosity questionnaire had been

completed by the twins, Ss read directions which introduced them to the subject of emotional descriptions and defined the task for them. They then turned to the first emotion and described a situation which made them feel *delighted*, for example. Following this description, they were asked to describe how they felt during the experience of Delight. Prior to their actually beginning to write, any questions they had were dealt with, and they were encouraged to try to recapture and to describe their feelings as vividly and as richly as they could. The distinction between the situational descriptions and the descriptions of their feelings was stressed at this time. The eight emotions were presented in random order, the questionnaires were randomly numbered, and all the protocols were content analyzed without knowledge of which group S was in.

The second part of the questionnaire consisted of several brief subtests. The first of these was a 20-item vocabulary test (Thorndike & Gallup, 1944), included because one would expect verbal intelligence to be a factor influencing the language of emotional experience, and any differences appearing along this dimension among the groups would have to be taken into account when interpreting the data.

An abbreviated (50-item) adjective checklist was the next subtest. The Ss were asked to describe themselves and then at the end of the questionnaire to describe their co-twin or sibling, using only 15 adjectives. Within each of the three groups, half the pairs described themselves first and half described their partner first. Two measures based on these checklists were obtained: (a) *Assumed Similarity*, defined as the number of adjectives checked in common on the self-descriptions and the descriptions of the twin or sibling, and (b) *Actual Similarity*, the number of adjectives checked in common on the self-descriptions of each member of a pair.

The first 10 questions from the Comprehension subtest of the Wechsler Intelligence Scale for Children (WISC) appeared next in the questionnaire. These were included with the idea of looking at the similarity of the language of twins in response to a standard stimulus in a "nonemotional" context.

Finally, the twins completed a checklist concerning the activities they did together, such as belonging to the same clubs, going to parties, etc. This was a rough measure of the degree of similarity of environment and was included since similarity of environment might affect the similarity of language. It consisted of 12 true-false items, scored in the direction of independence (false). The score was the total number of *false* responses made by a pair.

Administration

The questionnaires were administered in Ss' homes. Nearly always, those who knew each other were tested together in small groups. In the few cases where this was impossible to arrange, the exhortation that they not discuss the questionnaire seems to have been effective, since one S volunteered that he had tried to find out what the study was about, but that his friend, an earlier S, had refused to tell him.

However, pair members were always tested together. The time taken to complete the questionnaire ranged 1½–3 hr.

Analysis of Emotional Descriptions

Once the emotional experiences were obtained, they were analyzed according to a modified version of Farmer's (1967) system of classification. The system divided all possible responses into a number of categories, the most important of which were:

1. *Physiological responses.* These references to awareness of immediate stimulations from the body included 20 major categories consisting of 41 subcategories, many reflecting varying levels of intensity of feeling (high, medium, low) or distinguishing between sensations felt globally and those experienced in a specific part of the body. Some of the major categories were Activation Level (AL), Discomfort (D), Tension (T), Weight (W), and Cardiovascular (C). Examples of responses falling under these categories follow, and the appropriate subcategory is also indicated: "I was in a frenzy" (AL, high); "I feel tired" (AL, low); "I ache all over" (D, general); "My jaw feels stiff and tight" (T, specific); "My head felt like a ton" (W, heavy); "My heart starts to pound" (C, intensification).

2. *Functioning.* This involved references to emotional, motor, cognitive, sensory, and speech functioning while experiencing the emotion. These five categories were subdivided virtually the same way: Intensification, Enhancement, Depression or Slowing Down, and Dysfunction. Examples include the following: "I can't think straight" (Cognitive Dysfunction); "I feel graceful" (Motor Enhancement); "My senses seemed dulled" (Sensory Depression); "I was higher than high on liquor" (Emotional Intensification); "I can't talk" (Speech Dysfunction).

3. *Expressive behavior.* This category included references to laughing, smiling, crying, gesturing, and noninterpersonal shouting (e.g., "I laugh a lot"; "I felt like crying").

4. *Interpersonal relations.* References in this category referred to a relationship between S and another person or animal, including such major subcategories as Moving Toward, Moving Away, and Moving Against another. Moving Toward subcategories ranged from General Positive Feeling ("I felt friendly") to Wish for Proximity or Physical Contact ("I wanted to hug him"), with categories in between these extremes covering various degrees of closeness. Similar gradations of feeling were reflected in the categorization of responses falling under the category, Moving Away. These ranged from Lack of Identification ("I was revolted by them") and Avoidance of Interaction ("I did not want to talk to anyone") to Physical Withdrawal ("I felt like hiding") and Isolation ("I felt abandoned"). The category Moving Against included such subcategories as General Negativism ("I was fed up with them"), Passive Negativism ("I pretended not to hear her"), Verbal Aggression ("I gave an angry

retort"), and Physical Attack ("I punched him hard").

5. *Attitudes.* These references to a system of beliefs or feelings usually bound to an object and involving an evaluation of it included general attitudes as well as attitudes toward the self. Positive feelings, negative feelings, and hostile or self-destructive feelings were among those distinguished.

6. *Change.* Here references were to wishes either to resist or hasten change.

7. *Control.* References in this category were to various degrees of control, ranging from a sense of tight control to one of freedom ("I keep a tight grip on myself"; "I feel happy-go-lucky").

8. *Time.* These references were to a sense of the passage of time, either of its speeding by or of its slowing down.

The extent to which these categories were subdivided has been described at length (Allerand, 1967). In all, there were 127 different categories. For computational purposes no distinctions were made among them; they all counted equally in the totaling.

Since Farmer's (1967) categories were developed from material gathered from children, some changes were made to do justice to the more elaborated responses of the present adolescent sample. Additional subcategories were added when a given category could logically be subdivided further. In this study, for example, the category Time was divided into (a) Sense of Speeding of Time, and (b) Sense of Slowing of Time. This older sample, as it turned out, did make fairly frequent references to each of these subcategories. Very few of the categories were derived entirely from the present sample. One that was, however, was Color, which had not been anticipated, but was included to cover such rare responses as "Everything was yellow," and "I feel purple," remarks made by identical twin girls with reference to two different emotions.

Reliability

Fifty descriptions of emotional experiences were used in the reliability check. Two judges coded these, making 460 judgments in all and agreeing 89% of the time.

Similarity Evaluation

When all the data had been coded, the protocols were identified, and the intrapair similarity was evaluated. Each time both members of a pair received the same code, that is, used the same subcategory in their descriptions of the same emotion, a "hit" was scored. For example, if S said that when he felt delighted, he was "ready to explode," and if his twin reported feeling "ready to burst with delight," a "hit" was recorded, since both these responses fall under the same category (Pressure). If these two responses had referred to two different emotions, Anger and Delight, for example, then a hit would not have been counted.

To be considered a hit, responses had to refer to very similar experiences. Some of the physiological

responses, for example, were subdivided to distinguish between global and local bodily feelings. Thus in describing Fear, one twin might say he felt tense all over, while his co-twin felt his stomach tied up in knots during his experience of Fear. Though both report feeling tension when afraid, a hit would not be scored since their responses fall into two distinct categories (Tension, general and Tension, specific).

Another example of a hit would be if S wrote that when he felt angry, he had an impulse to "shout insults at someone," and if his twin described "speaking sarcastically" when he was angry. In this case, both responses would be scored as Verbal Aggression. However, if the second twin had said he felt like throwing something at someone, a hit would not have been recorded since this response constitutes a separate category, Desired Physical Attack.

Emotional Similarity Index

When all the data had been analyzed for "hits," that is, for identical subcategories within the same emotion, the average number of categories used across all the emotions was computed for each pair. The proportion of hits to mean categories was then calculated. For example, if Twin A and Twin B used 25 and 35 categories in their respective descriptions, and if they used the same categories in describing the same emotion 10 times (10 hits), then their overall proportion of similarity would be $10/30$ or .33. This proportion is referred to as the Emotional Similarity Index (ESI).

RESULTS

Among the MZ twins the mean ESI (.201, $\sigma = .08$) was found to be considerably higher than that of either DZ twins (.147, $\sigma = .06$) or siblings (.131, $\sigma = .06$). Since the ESI was computed as a proportion of categories used in the descriptions of emotions, a square-root transformation of the scores was performed to insure the homogeneity of the variance among the three groups. These transformed scores were used to compute the analysis of variance summarized in Table 2, which indicates a sig-

TABLE 2
ANALYSIS OF VARIANCE OF THE EMOTIONAL
SIMILARITY INDEX
(TRANSFORMED SCORES)

Source	df	MS	F
Groups (MZ, DZ, Sib)	2	.0569	5.63*
Sex	1	.0320	3.17
Interaction	2	.0094	.93
Error	66	.0101	
Total	71		

* $p < .01$.

nificant difference ($F = 5.63, p < .01$) among the groups. Neither the difference between pairs of male and female Ss nor the interaction Sex \times Group identity was significant.

In view of the significant differences found in the analysis of variance, two t tests were computed, one to evaluate the difference between the MZs and DZs, and the other to compare the DZ twins and Sibs. Since these t tests were done after the analysis of variance, a significance level of .02 (two-tailed) was used instead of a level of .05. This more conservative significance level was adopted in the spirit of the multiple-comparison procedure proposed by Dunn (1961). The difference between the transformed means of the MZ and DZ twins (.44 and .37, respectively) was significant beyond the .02 level ($t = 2.43$), while the difference between the means of the DZ twins and the Sibs (.37 and .35, respectively) was not statistically significant ($t = .69, .2 < p < .5$). Therefore, the data supported the hypothesis that MZ twins are more similar in their descriptions of emotional experiences than either DZs or Sibs. However, the prediction of greater similarity among DZs as compared with Sibs was not sustained.

Socioeconomic Status

Since no relationship was found between the estimated social class rating and ESI ($r = -.07$), the observed difference in socioeconomic level between the MZs and the others need not be considered as having any significant bearing on the finding that heredity does play a part in determining the report of the subjective experience of emotions.

Richness of Description

Two measures of the richness of descriptions were made. The first compared the

TABLE 3

RICHNESS OF DESCRIPTION: NUMBER OF DIFFERENT CATEGORIES USED PER PAIR

Group	MZ		DZ		Sib	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Males	30.3	8.9	33.1	9.0	37.0	9.3
Females	41.6	9.7	44.3	13.9	37.9	7.1
<i>t</i>	2.87**		2.32*			

* $p < .05$.

** $p < .01$.

total number of categories used over all emotions for each S with that of his co-twin or sibling. The correlations, .69 (MZ), .75 (DZ), and .43 (Sib), indicated that there was a tendency for twins to use the same number of categories in their emotional descriptions. Both correlations for the twins were significantly different from zero ($p < .02$, two-tailed test).

A more refined measure of richness of description was made by computing the total number of different categories used by each pair. Table 3 shows the means and standard deviations of these scores according to sex and group identity. Significant sex differences were found only among the twins. MZ differences were significant at the .01 level, while DZ differences reached the .05 level of significance. Group differences were not significant.

Similarity of Activities

Analysis of the Twin Activity Checklist (one score per pair equal to the total number of items answered by each in the direction of greater independence) revealed strong support for previous findings that MZ twins form closer bonds of attachment and exhibit greater intrapair similarity with respect to habits and activities (Smith, 1965). The difference between the mean of the MZs (8.54, $\sigma = 4.03$) and the mean of the DZs (12.29, $\sigma = 4.42$) was significant at the .01 level (two-tailed test), indicating more independent activities among the DZ twins. The correlation between the Activity score and the ESI for each pair was computed, since the greater similarity of the emotional experience of MZ twins might be attributable to their spending more time together. The results (MZ, $r = .01$; DZ, $r = .05$; MZ + DZ, $r = -.11$) ruled out this possibility.

Actual and Assumed Similarity of Personality

It is clear from Table 4 that MZ twins strongly assume their twin's personality and their own to be similar ($t = 4.17, p < .002$). DZ twins and Sibs did not differ in this respect. However, there were no differences among the groups with respect to Actual Similarity, nor was there a significant difference between the MZ twins' mean Assumed Similarity score and their Actual Similarity score.

The correlations between the ESI and the Actual Similarity scores were computed to determine if those pairs who describe themselves most similarly also tended to experience emotions most similarly, but no relationship was found (MZ, $r = .07$; DZ, $r = -.07$; Sib, $r = .05$).

WISC Comprehension Subtest

Finally, in an attempt to see if the language habits of MZ twins are more alike regardless of what they are talking about, the number of identical words appearing in the twins' answers to the first 10 WISC Comprehension questions were counted. Twin A's answer to each question was compared with Twin B's answer to the same question, and the number of identical words (excluding *a*, *an*, and *the*) appearing in both answers was tabulated, and so on for all 10 questions for each pair of twins. The mean number of identical words per pair was 24.2 (MZ) and 25.5 (DZ) with standard deviations of 11.5 and 13.8, respectively. Thus it can be concluded that in general, MZ twins did not tend to use the same words more often than DZs, and therefore the discovered similarity of their language of experience appeared to be independent of their general language habits.

DISCUSSION

The results of this study lend support to the idea that heredity in part determines reported phenomenological experience. While the overall effect was relatively small, it is important to note that the ESI increased considerably with increased genetic similarity: the MZs' mean ESI (20.1%) represented an increase of 36% when compared with the DZs' mean (14.7%) and an increase of 53% when compared with the mean of the Sibs (13.1%). This was an average increase of about 45% in going from a group with an average of 50% of their genes in common (DZs and Sibs) to one with all of their genes the same (MZs). Ideally, of course, this study would have included an additional control group of unrelated pairs matched for age and sex.

This finding is consistent with the mounting evidence that affective experiences and behavior are in part genetically determined

TABLE 4
ASSUMED SIMILARITY AND ACTUAL SIMILARITY SCORES

Score	MZ		DZ		Sib	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Assumed	9.0*	2.2	7.1	2.4	6.7	2.5
Actual	8.0	2.2	7.0	1.7	6.4	1.9

Note.—*N* = 15 of 50 items.

* $p < .002$, $t = 4.17$, in comparison with DZ.

(Freedman, 1965; Gottesman, 1963). In view of the fact that social introversion (Gottesman, 1963), sociability (Scarr, 1965), and extraversion-introversion (Eysenck, 1956; Shields, 1962) have all been found to have a genetic basis, it seems reasonable to suppose that persons who differ with respect to this general trait of involvement with others experience emotions differently. The highly sociable person, for example, might be more likely to report changes in his relationships with others during different emotional experiences, while the introverted, less sociable person might be more likely to focus on changes within himself. Introversion, furthermore, has been found to be associated with strong internal inhibition (McConaghy, 1962), and persons with a great capacity to delay gratification have been found to have a longer, future-time perspective than those whose internal inhibitions are weaker (Kahn, 1967). Consequently, one might expect the subjective experience of MZ twins to reflect a more similar sense of time. There is also evidence that extraverts tend to outbursts (McConaghy, 1962). Again, these characteristics might well be reflected in descriptions of emotional experiences.

A frequent criticism of the classical twin study is that the finding of greater intrapair similarity among MZ twins may be attributable to greater environmental similarity rather than to heredity. The present study has found evidence which does not support this view. It remains to be shown, however, that those aspects of the MZs' environment which are more similar have a significant influence on their experience of emotions. Attributing MZ similarities primarily to environmental effects is too simplistic a solution.

The problem of environmental similarity is

indeed a difficult one with which to deal, and critics of the twin method often suggest that parents of MZs create similarities between them by the very fact that they consider them to be identical, while the parents of DZ twins in considering them to be different help to create differences.

A critical test of environmental bias is possible by examining the behavior and attitudes of parents who are wrong about the zygosity of their twins. Scarr (1968) has done just this, and by comparing them with those of parents of correctly identified twins, she was able to separate their reactions to similarities and differences based on genetic relatedness from those arising from their belief that their twins should or should not be similar.

Though Scarr's sample was admittedly small ($N = 12$), strong trends were found in the data suggesting that genetic relatedness of twins determined the similarity of treatment by parents, rather than the parents' belief as to the zygosity of their children. That is, MZ twins incorrectly considered to be DZ were treated more like correctly identified MZ pairs by their mothers, while DZ twins misclassified MZ were treated more like correctly identified DZ pairs.

Some investigators have found evidence that special differences exist in the environment of identical twins (Eysenck & Prell, 1951; Misbach & Stromberg, 1941). The presence of any of these factors can lead only to the underestimation of the effect of heredity.

Similarity of reported emotional experience is closely linked to similarity of expressiveness, and while it might be thought that emotional expressiveness is primarily learned through parent-child interaction and that for this reason MZ twins are more similar with respect to their subjective experiences, Bronson (1966) has shown that this is not the case. She found that the level of emotional expressiveness in a child is not a function of parental attitude towards him, but depends on the kind of person the parent is. Therefore the ESIs may be considered to be independent of this trait.

Two phenomena may be involved in the finding that female twins gave significantly richer descriptions than male twins. The well-

known greater verbal facility of girls is precluded as an explanation, as is inhibition among the boys due to the female sex of the investigator, since among Sibs, there were no sex differences with respect to richness. In our present society there is considerable pressure on boys to repress or suppress their feelings. In many cases to be a man is to be emotionally controlled, and this tends to work toward decreased emotional sensitivity both toward oneself and others. Consequently, boys may not experience their feelings as freely as girls.

The data suggested that richness of emotional experience is also increased by having a same-age peer in whom one can confide, and that this experience is usually more common among girls than boys because of social pressures. Thus each female twin has an unusual opportunity to share her experiences, and in so doing, she may enlarge her range of response not only through the addition of some of her twin's ways of reacting to her own repertoire, but also through the making explicit of some of her heretofore unlabeled or unrecognized feelings.

REFERENCES

- ALLERAND, A.-M. Remembrance of feelings past: A study of phenomenological genetics. Unpublished doctoral dissertation, Teachers College, Columbia University, 1967.
- ARNOLD, M. Physiological differentiation of emotional states. *Psychological Review*, 1945, 52, 35-48.
- AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, 1953, 15, 433-442.
- BRONSON, W. Early antecedents of emotional expressiveness and control. *Child Development*, 1966, 37, 793-810.
- DUFFY, E. Level of muscular tension as an aspect of personality. *Journal of General Psychology*, 1946, 35, 161-171.
- DUFFY, E. *Activation and behavior*. New York: Wiley, 1962.
- DUNN, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, 1961, 56, 52-64.
- EYSENCK, H. J. The inheritance of extraversion-introversion. *Acta Psychologica*, 1956, 12, 95-110.
- EYSENCK, H. J., & PRELL, D. B. The inheritance of neuroticism: An experimental study. *Journal of Mental Science*, 1951, 97, 441-465.
- FARMER, C. Words and feelings: A developmental study of the language of emotion in children. Unpublished doctoral dissertation, Teachers College, Columbia University, 1967.

- FREEDMAN, D. An ethological approach to the genetical study of human behavior. In S. Vandenberg (Ed.), *Methods and goals in human behavior genetics*. New York: Academic Press, 1965.
- GOTTESMAN, I. Heritability of personality: A demonstration. *Psychological Monographs*, 1963, 77(9, Whole No. 572).
- GRAHAM, D. T. Cutaneous vascular reactions in Raynaud's disease and in states of hostility, anxiety, and depression. *Psychosomatic Medicine*, 1955, 17, 200-207.
- HUNT, J. McV., COLE, M. W., & REIS, E. Situational cues distinguishing anger, fear, and sorrow. *American Journal of Psychology*, 1958, 71, 136-151.
- JOST, H., & SONTAG, L. W. The genetic factor in autonomic nervous system function. *Psychosomatic Medicine*, 1944, 6, 308-310.
- KAHN, P. Time span and Rorschach human responses. *Journal of Consulting Psychology*, 1967, 31, 92-93.
- LACEY, J. I., & VAN LEHN, R. Differential emphasis in somatic response to stress. *Psychosomatic Medicine*, 1952, 14, 71-81.
- MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanism in psychiatric patients under stress. *Psychosomatic Medicine*, 1949, 11, 25-29.
- MC CONAGHY, N. The inhibitory index in relation to extraversion-introversion. *American Journal of Psychiatry*, 1962, 119, 527-533.
- MINER, J. *Intelligence in the United States*. New York: Springer, 1957.
- MISBACH, L., & STROMBERG, R. N. Nonseparation as a source of dissimilarities between monozygotic twins: A case report. *Journal of Genetic Psychology*, 1941, 59, 249-257.
- MORDKOFF, A. M. The relationship between psychological and physiological response to stress. *Psychosomatic Medicine*, 1964, 26, 135-150.
- NICHOLS, R. C., & BILBRO, W. C., JR. The diagnosis of twin zygosity. *Acta Genetica*, 1966, 16, 265-275.
- RICHMOND, J. M., & LUSTMAN, S. L. Autonomic functions in the neonate. I. Implications for psychosomatic theory. *Psychosomatic Medicine*, 1955, 17, 269-275.
- RUESCH, J., & FINESINGER, J. E. The relation between electromyographic measurements and subjective reports of muscular relaxation. *Psychosomatic Medicine*, 1943, 5, 134-138.
- SCARR, S. The inheritance of sociability. Paper presented at the annual meeting of the American Psychological Association, Chicago, September 1965.
- SCARR, S. Genetic factors in activity motivation. *Child Development*, 1966, 37, 663-673.
- SCARR, S. Environmental bias in twin studies. *Eugenics Quarterly*, 1968, 15, 34-40.
- SCHACHTER, J. Pain, fear, and anger in hypertensives and normotensives. *Psychosomatic Medicine*, 1957, 19, 17-29.
- SCHACHTER, S. The interaction of cognitive and physiological determinants of emotional states. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964.
- SCHAFER, H. R. Activity level as a constitutional determinant of infantile reaction to deprivation. *Child Development*, 1966, 37, 595-602.
- SHAGASS, C., & KERENYI, A. B. Neurophysiologic studies of personality. *Journal of Nervous and Mental Disease*, 1958, 126, 141-147.
- SHIELDS, J. *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press, 1962.
- SMITH, R. T. A comparison of socioeconomic factors in monozygotic and dizygotic twins, testing an assumption. In S. Vandenberg (Ed.), *Methods and goals in human behavior genetics*. New York: Academic Press, 1965.
- THERON, P. A. Peripheral vasomotor reactions as indices of basic emotional tension and lability. *Psychosomatic Medicine*, 1948, 10, 335-346.
- THORNDIKE, R. L., & GALLUP, G. Verbal intelligence of the American adult. *Journal of General Psychology*, 1944, 30, 75-85.
- WARNER, W. L., MEEKER, M., & EELIS, K. *Social class in America*. Chicago: Science Research Associates, 1949.

(Received December 19, 1967)

EFFECT OF CONTEXT AND STRENGTH OF ASSOCIATION ON SCHIZOPHRENIC VERBAL BEHAVIOR

C. WILLIAM DECKNER

AND

RICHARD L. BLANTON

*Veterans' Administration Hospital,
Murfreesboro, Tennessee*

Vanderbilt University

Rate of decline of performance as a function of progressively weakened context was found to be no greater for schizophrenics than for normals. The variable of strength of association, however, was found to differentiate the verbal behavior of schizophrenics from that of normals; schizophrenics performed at a level comparable to normals when rewarded for choosing the strong associations of stimulus words but were significantly inferior to normals when rewarded for choosing weak associations.

Chapman, Chapman, and Miller (1964) presented evidence that the verbal behavior of schizophrenics is characterized by an excessive reliance on the strongest normal association to a word and a consequent neglect of a word's weaker associations. They also suggest that the extent to which this tendency differentiates schizophrenics from normals is determined by the strength of the contextual cues which serve to guide the associative response. When contextual cues which indicate the appropriate associative response are very strong, schizophrenic performance should closely approximate that of normals; when such contextual cues are reduced, schizophrenic performance should deteriorate relative to that of normals.

The two experiments reported here were attempts to examine the generality of the theory by using methods to define context and strength of association which are somewhat different from those of Chapman et al.

EXPERIMENT I

The first experiment examined the hypothesis that schizophrenics are unable to use weak contextual cues to the same degree as normal Ss.

Method

Materials. The method used to define context was cloze procedure (Taylor, 1953). Words were systematically deleted from a passage, and the remaining context provided cues to the deleted words.

¹ Requests for reprints should be sent to C. William Deckner, Department of Psychiatry, Vanderbilt University, Nashville, Tennessee 37203.

Either every tenth, eighth, or fifth word was deleted from three passages of first-grade difficulty and three passages of third-grade difficulty (*Webster's Practice Readers*, Grover & Bayles, 1942). Thus, a test unit was comprised of six different passages of three frequencies of deletion and two levels of difficulty. No passage appeared more than once in a test unit, and the order of the passages was counterbalanced.

Subjects. A group of 36 male schizophrenics divided into two groups of 18 on the basis of their scores on the Phillips Scale of Premorbid Adjustment (Phillips, 1953) and a control group of 18 male general medical patients were Ss. All Ss were Veterans' Administration Hospital patients. The good premorbid and poor premorbid groups, respectively, had a mean of 7.09 and 9.93 yr. since first admission. Both groups were variable with regard to subclassification of schizophrenia. The mean IQ scores obtained on the Peabody Picture Vocabulary Test (Dunn, 1954) by the normals, good premorbid schizophrenics, and poor premorbid schizophrenics were 96.78, 101.33, and 99.05, respectively. Their mean ages were 37.83, 40.89, and 40.39, respectively. Patients who were over 50 yr. of age, who had less than a sixth-grade education, or who were unable to read at the fifth-grade level were not used. No schizophrenic Ss employed had at any time been given any other psychiatric diagnosis. No normal Ss employed had been given psychiatric diagnosis of any kind.

Procedure. To begin, Ss were given a test of reading ability, shown a sample passage, and given a warm-up passage to complete. They were instructed: "There are words missing from these short stories. Please make your best guess as to what the words should be and then write them back in the blank spaces. Write only one word in each blank space." Following the warm-up passage, *E* answered procedural questions and instructed *S* to complete the test passages.

The words supplied by Ss were given appropriate scores of 2, 1, or 0 points by two college students who were trained on the judging task and established to have a high interjudge reliability ($r = .85$). Neither judge was informed as to the

purpose of the study, and identifying data on the response sheets were in code. The sum of the judges' ratings was used as the measure of the dependent variable for each deletion.

Prediction. Schizophrenic performance was predicted to be more impaired than that of controls by increases in contextual deletion. Schizophrenic performance, therefore, was expected to approximate normal performance most closely with every tenth word deleted (minimal deletion), and be most inferior to normal performance with every fifth word deleted (maximal deletion).

Results

Table 1 is the summary table of the analysis of variance of the total cloze scores.

Most pertinent to the authors' prediction is that while the main effect of frequency of deletion reached significance ($p < .05$), this variable did not differentially influence the performance of the three groups of Ss. The Groups \times Deletions interaction was not significant ($p > .15$) which showed that the rate of decline as a function of frequency of deletion was probably no greater for schizophrenics than for normals.

Of less importance, the overall level of performance of the schizophrenics was inferior. As is shown, the main effects of the groups variable ($p < .05$) and the level of difficulty ($p < .01$) were significant. The only interaction which closely approached significance was Level of Difficulty \times Frequency of Deletion ($p = .06$).

Perhaps a purer measure of the ability to use weak contextual cues is the number of 2-point (maximum credit) words supplied. (One-point words were excluded in this analysis.) Using this dependent variable, the main effect of frequency of deletion was found to be highly significant ($p < .01$), but again the performance of the three groups was not differentially affected (Groups \times Deletions interaction $p > .15$). Again the schizophrenics were generally inferior to the normals ($p < .05$), and the main effect of level of difficulty was significant ($p < .01$). No interactions reached statistical significance.

EXPERIMENT II

The second experiment was an attempt to demonstrate the schizophrenic's hypothesized exaggerated reliance on the strongest association evoked by a word and relative neglect of

TABLE 1
ANALYSIS OF VARIANCE OF CLOZE TASK PERFORMANCE
AS A FUNCTION OF LEVEL OF DIFFICULTY
AND AMOUNT OF DELETION

Source of variation	df	MS	F
Between Ss	53		
Normals, good PM and poor PM (A)	2	900.54	3.27*
Ss within groups	51	275.74	
Within Ss	270		
Level of difficulty (B)	1	1,190.25	60.12**
A \times B	2	34.48	1.74
B \times Ss within groups	51	19.80	
Frequency of deletion (C)	2	56.92	3.11*
A \times C	4	31.68	1.73
C \times Ss within groups	102	18.28	
B \times C	2	39.58	2.96
A \times B \times C	4	9.28	
B \times C \times Ss within groups	102	13.26	

* $p < .05$.

** $p < .01$.

a word's weaker associations. A concept-form task was constructed for this purpose.

Method

Materials. Materials consisted of the Minnesota Word Association Norms (Jenkins & Palermo, 1963) and two decks of 3×5 in. cards on which stimulus and response words were printed.

Each deck consisted of 55 pairs of cards. On the first card of each pair was a stimulus word shown by the Minnesota norms to elicit a response with high commonality (e.g., NEEDLE). Immediately following and completing the pair was a card with two response words, one of which accounts for the high commonality (e.g., "thread"). The other response word was one which the norms show is given with some frequency as an association to the stimulus word (e.g., "nail") but much less frequently than the alternative previously described. Strength of association, therefore, was defined in terms of cultural frequency of association (cf. Mednick & Devito, 1958). The two response words were the same part of speech, either both concrete or both abstract, and randomly positioned on the card.

Subjects. The Ss were the same as those used in Experiment I.

Procedure. Each S was seated across a table from E. With illustrations, the following instructions were given:

All you have to do is look at a card with a single word on it. Then look at a second card which has two words on it and choose one of these two words by saying it aloud. One of the words is the nickel word. Each time you choose the nickel word, I will place a nickel on the table beside you. All the money you win will be yours to keep.

There was no time limit and S was permitted to progress through the cards at his own rate. Every effort was made to have Ss feel as little pressure as possible.

TABLE 2
ANALYSIS OF VARIANCE OF TRIALS TO CRITERION ON
CONCEPT FORMATION TASK WITH TWO
LEVELS OF ASSOCIATIVE STRENGTH

Source of variation	df	MS	F
Between Ss	53		
Normals, good PM, and poor PM (A)	2	1,322.01	8.69*
Order of shift (B)	1	293.37	1.93
A \times B	2	71.29	
Ss within Groups A and B	48	152.15	
Within Ss	54		
Strength of association (C)	1	13,511.70	106.57*
A \times C	2	817.68	6.45*
B \times C	1	0.33	
A \times B \times C	2	395.36	3.12
C \times Ss with Groups A and B	48	126.79	

* $p < .01$.

For half of both the schizophrenic and normal groups the "nickel word" in the first deck of cards was the strong association to the stimulus word. For the other half groups the reward was given for choosing the weak association to the stimulus word.

After each S had either reached the criterion of 9 out of 11 correct choices or had progressed through the entire deck of 55 sets, E put aside the first deck and said: "That's fine. Now see if you can guess the nickel word with these cards." With the introduction of the second deck, a shift was required; Ss previously reinforced for choosing the strong associations were now reinforced for choosing the weak associations, and vice versa.

As no stimulus word or response word was repeated throughout the two decks, the task was not one of remembering which response was rewarded on a previous trial. In order to reach criterion, an S was

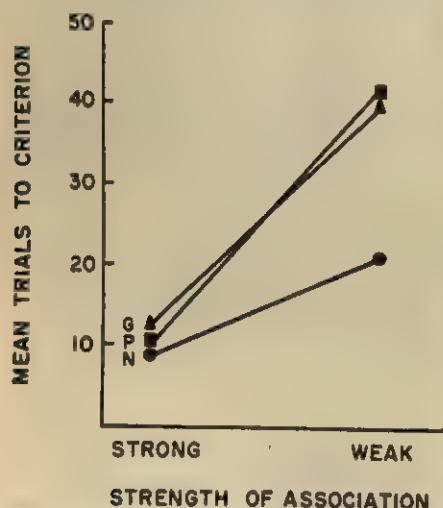


FIG. 1. Mean trials to criterion as a function of strength of association in good and poor premorbid schizophrenic and normal Ss. (G, triangles: good premorbid schizophrenics; P, squares: poor premorbid schizophrenics; N, circles: normals.)

assumed to have attained the concept of associative strength or at least some formulation related to it.

Predictions. No difference was predicted between the schizophrenic and normal groups when reinforcement was contingent upon choosing strong associations, while a significant difference was predicted between the groups when reinforcement was contingent upon choosing weak associations.

Results

Table 2 is the summary of the analysis of variance on the number of trials (up to 55) required to reach criterion. The main effect of the strength of association variable was significant, indicating that all Ss performed more poorly when weak rather than strong associations were reinforced ($p < .01$). Furthermore, as predicted, the interaction Groups \times Strength of Association ($p < .01$) was significant.

To clarify the source of the interaction, the differences among groups were analyzed separately at each level of strength of association by simple analysis of variance. When reinforced for responding to strong associations, Ss were not different. However, group differences did occur ($p < .01$) when weak associations were reinforced (see Figure 1). A Neuman-Keuls test indicated that the schizophrenic groups, while not different from each other, performed more poorly than the normal group ($p < .01$).

Of less importance, the groups main effect was significant ($p < .01$), indicating the normals to be superior to both schizophrenic groups. However, the generality of this finding is limited by the interaction finding already described. The order of shift main effect and the other interactions did not reach significance.

DISCUSSION

Chapman, Chapman, and Miller (1964) suggest that the "meaning responses" elicited by words are only one form of associative response. In Experiment II it was found that schizophrenics manifest the predicted behaviors not only in response to the stronger and weaker meanings of words, but also in response to the stronger and weaker associations elicited by words of a variety of types. The schizophrenics showed inferior performance relative to normals when rewarded for choosing the weak associations but performed at a level comparable to normals when re-

warded for choosing the strong associations. This appears to provide a broader base of support for Chapman's hypothesis.

Contrary to prediction, in Experiment I it was found that the performances of schizophrenics and normals decline uniformly with increases in frequency of deletion. This suggests that of the two variables specified by Chapman et al., schizophrenic verbal behavior may be primarily determined by the strength of association variable. The variable of context may not differentiate schizophrenics from normals at all, or it may operate only in interaction with strength of association. Since the two variables were measured independently in the two experiments such an interaction could not occur and the effect of context would not be demonstrated.

It is also possible that the present definition of context is too general, and that "contextual constraint" is provided, in Chapman's sense, only by certain classes of words; for example, adjectives and nouns might constrain meaning in a different way from prepositions or conjunctions. Since Chapman's definition of constraint does not specify such factors, they remain a matter for further study.

The inconsistency of the findings with the cloze task with those of Chapman also suggests the relevance of attentional dysfunction theory in schizophrenia (Silverman, 1964; Venables, 1964). Chapman's multiple-choice tasks emphasize the ability to use critical cues which are spatially separated from the point of associative choice. For example, "farmer" and "cows" are the critical contextual cues in the following item.

The farmer has too many cows and must buy a new *pen*.

The word *pen* in this sentence means:

1. Pick-up truck
2. Writing implement
3. Fenced enclosure

In contrast, the critical cues in a cloze task are very likely to be in close proximity to the

point of associative choice; that is, the words adjacent to the blanks are those most likely to be decisive and also those immediately present in the stimulus field (see Cromwell & Dokecki, 1968). An impairment of attention in schizophrenia would make this difference between the two experimental tasks critical for differences in performance; the same results could not necessarily be expected from the two procedures. However, the discrepancy which occurs points to the need for greater specification of the types of contextual constraints which differentiate the verbal behavior of schizophrenics from that of normals.

REFERENCES

- CHAPMAN, L. J., CHAPMAN, J. P., & MILLER, G. A. A theory of verbal behavior in schizophrenia. In B. Maher (Ed.), *Progress in experimental personality research*, Vol. 1. New York: Academic Press, 1964.
- CROMWELL, R. L., & DOKECKI, P. R. Schizophrenic language: A disattention interpretation. In S. Rosenberg & J. H. Koplin (Eds.), *Developments in applied psycholinguistics research*. New York: Macmillan, 1968.
- DUNN, L. M. *Peabody Picture Vocabulary Test*. Minneapolis, Minn.: American Guidance Service, 1954.
- GROVER, C. C., & BAYLES, C. R. *Webster's practice readers*. New York: Webster, 1942.
- JENKINS, J. J., & PALERMO, D. S. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.
- MEDNICK, S. A., & DEVITO, S. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, 55, 316-327.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 352-379.
- TAYLOR, W. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, 30, 415-433.
- VENABLES, P. H. Input dysfunction in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 1. New York: Academic Press, 1964.

(Received December 21, 1967)

COGNITIVE LEVEL AS A FUNCTION OF DEFENSIVE ISOLATION¹

LILA LOWENHERZ AND MELVIN FEFFER²

Yeshiva University

In previous studies the cognitive structuring of interpersonal content has been assessed by means of a role-taking task (RTT) which requires the individual to tell a standard Thematic Apperception Test (TAT) story to an ambiguous picture and then to retell the story from the perspective of each of his TAT figures. The present study evaluated the possibility that a failure to coordinate perspectives on the RTT occurs when the TAT figures represent defensively isolated aspects of the self-structure. The Ss were given the RTT under two different conditions. Under one condition, Ss were assigned defensively isolated attributes, asked to tell a story built around TAT figures endowed with these attributes, and then required to take the perspectives of these figures in their role taking. The other condition differed in that the assigned attributes were not defensively isolated. The Ss showed more difficulty in coordinating perspectives under the defensive isolation condition than under the nonisolated condition.

Piaget (1950) has characterized cognitive development as a reorganization of behavior in which focusing sequentially upon single aspects of the experiential field becomes subordinated to focusing upon a number of aspects of the field in relation to one another. Exemplifying this conception is Piaget's (1950) classic study of the child's judgment of quantity in which an invariant amount of liquid is poured successively into containers of different shapes. Piaget attributed the younger child's lack of conservation to his focusing upon one dimension of the situation at a time, for example, attending first to changes in the height of the liquid while neglecting compensatory changes in the circumference of the container and then shifting attention to circumference while neglecting height. By the same token, he viewed the older child's attainment of conservation as a function of his ability to coordinate changes along both dimensions, that is, to consider changes in height in relation to changes in

circumference. His position would suggest that a defining characteristic of "sequential" decentering in contrast to "relational" decentering is isolation or lack of contact between relevant aspects of experience.

In general, the decentering concept has been restricted to an interpretation of impersonal cognitive functioning. An extension of the decentering concept to the interpersonal area, however, has been suggested (Feffer, 1967) and evaluated empirically by means of a projective role-taking task (RTT). In the RTT, S is first required to tell a standard Thematic Apperception Test (TAT) story to an ambiguous drawing. He is then asked to retell the story from the viewpoint of each of the figures in his story. Decentering level has been measured in terms of the degree to which S is able to focus upon his initial story from the particular perspective of each of his figures while at the same time maintaining continuity between his different versions of the story. This measure of the ability to coordinate different perspectives has been found to be associated with decentering activity on impersonal conservation and part-whole tasks (Candell, 1965; Feffer & Gourevitch, 1960), chronological age (Candell, 1965; Feffer & Gourevitch, 1960; Wolfe, 1963), and effectiveness of social interaction (Feffer & Suchotliff, 1966).

¹ This article is derived from the first author's doctoral dissertation, completed at Yeshiva University in 1966 under the second author's supervision. The study was supported in part by Public Health Service Research Grant MH-07506 from the National Institute of Mental Health.

² Requests for reprints should be sent to Melvin Feffer, Department of Psychology, Yeshiva University, 55 Fifth Avenue, New York, New York 10003.

In the present study, the projective nature of the RTT itself becomes the focus of inquiry in that the possibility is raised that a dynamic factor is involved in S's ability to coordinate the different perspectives of his projected figures. This possibility is outlined in the following rationale: In constructing stories involving ambiguous TAT figures, individuals may typically express a variety of self-attributes, some of which have been defensively isolated as a means of avoiding the arousal of anxiety. In being required, in role taking, to assume the perspectives of the various figures in his stories, the individual is by that token required to act as the subject of these defensively isolated attributes. In a sense, then, he is forced to behave as though these attributes were syntonetic aspects of the self. Under these circumstances, a reactivated defense against a possible confrontation with the anxiety-provoking attribute could take the form of isolation or lack of coordination between the various perspectives involved. The present study was directed toward an hypothesis which follows from this rationale, namely, that a failure to coordinate perspectives will occur when the TAT figures represent defensively isolated aspects of the self-structure.

METHOD

The procedure in the first part of the study was directed toward determining for each individual, those attributes which were defensively isolated and those attributes which were not defensively isolated. In the second part of the study, Ss were given the RTT under two different conditions. Under one condition, Ss were assigned their defensively isolated attributes, asked to tell a story built around TAT figures endowed with these characteristics, and then required to take the perspectives of these figures in their role taking. The other condition differed in that the assigned attributes were not defensively isolated. A comparison of decentering level under these two conditions constituted the evaluation of the study's hypothesis. It was predicted that a greater degree of sequential decentering in role taking would be evidenced under the condition of defensive isolation than under the nonisolation condition.

Measure of Defensive Isolation

The projective assumption was basic to the measure of defensive isolation; that is, any personality characteristic which S ascribed to a TAT figure was assumed to be an attribute of the self. If, further,

S reacted to this attribute as being unacceptable and unlike the self, this was taken as an indication that the characteristic was defensively isolated from the rest of the self-structure.

The method for assessing defensive isolation, as so defined, followed a modified version of a procedure described in Pile, Mischel, and Bernstein (1959). A group of female second-year college students were shown seven stimulus cards projected onto a screen by an opaque projector. Five of the pictures were TAT cards (Numbers 4, 10, 12F, 13MF, and 18GF), and two were based upon the Symonds Picture Story Test (Numbers B3 and B8). The Ss were told that their stories were being collected as part of a larger study attempting to find a simple, reliable method of assessing creativity from the TAT. They were also told that selected Ss would be asked to return for further testing. Standard TAT instructions were used and 6 min. were allowed for each story.

The first author analyzed each TAT story for attitudinal content and, for each TAT figure, constructed one statement which best represented S's characterization of that figure. The criteria³ used in constructing statements have been described in detail in Lowenherz (1966) and were directed toward insuring that statements were similar in respect to clinically relevant variables. These criteria are summarized as follows:

1. The attributes of each TAT figure were drawn from that part of the story concerned with events taking place in the present.

Example of TAT content. I think he was faced with a great disappointment. . . . He seems to feel that he is a failure and is not worthy of his wife's love . . .

Derived Statement. He seems to feel that he is a failure and is not worthy of his wife's love. [The quality of disappointment takes place in the past and is not used in the construction of the statement.]

2. Statements about the TAT figure were drawn from the story directly, rather than from attributes which were ascribed to the figure by other characters in the story.

Example of TAT content. The husband was fired from his job. He's angry and barks at his wife, claiming that she was the one who caused his failure. In the picture, one sees the wife, who loves her husband dearly and doesn't like to see him feeling so defeated and worthless . . .

Derived statement. When angry, he blames others for his misfortunes. [This is chosen in preference to he feels defeated and worthless because the latter are feelings reflected through another character.]

³ There was agreement in 35 out of 40 instances (87%) with another person who independently constructed statements from the stories in terms of these criteria.

3. Attributes implicit in the story were used only when *S* did not explicitly and directly ascribe qualities to the TAT figures.

Example of TAT content. Laurie and Peter have only been married for a few months. Laurie has just returned from shopping and has found Peter lying at the bottom of the stairs. . . . He tells her that his back is broken and to call a doctor. . . . The doctor comes and Peter is taken to the hospital but his spinal cord was severed and he becomes an invalid and Laurie has to support them.

Derived statement. He is helpless and needs to be cared for. [No explicit quality has been ascribed to the male figure, hence an inference is made.]

4. Where appropriate, the statements were made reciprocal. Thus, "He is helpless and needs to be cared for" is complemented by the statement "She is helpful and supporting," for the female figure in the story aforementioned.

5. Statements were all changed to designate females. Thus, the statements illustrating Criteria 2 and 3 were changed to "When angry she blames others for her misfortunes" and "She is helpless and needs to be cared for."

Each *S* produced stories from which it was possible to construct statements characterizing between 6 and 12 TAT figures. For those *Ss* who produced less than 12 statements, "fillers" were randomly assigned from the total pool of statements so that there were 12 statements per *S*. The statements were typed on ditto stencils and run off; thus, although individually tailored, each packet of statements had the appearance of a group inventory.

About 5 wk. after the first session, *Ss* were again seen in a group and were told that they were going to be given a different type of task. Each *S* was given her individually tailored packet of statements and was asked to rank these from "most like me" to "least like me." In addition, each was asked to dichotomize an identical set of statements into the categories of "personally acceptable" and "not personally acceptable." There was no indication that any of the *Ss* saw any connection between the statements and the TAT which had been administered earlier.

An attribute was defined as defensively isolated if *S* considered it as "not personally acceptable" and had placed it in either of the two most extreme rankings in the "less like me" direction. A non-isolated attribute was defined as one which *S* considered "personally acceptable" and which she had placed above the midpoint of her rankings in the "more like me" direction. Where possible, a pair of isolated and a pair of nonisolated attributes were selected for each *S*. Each attribute of the pair was drawn from a separate TAT card. Attributes which served as fillers for a given *S* were not defined as either isolated or nonisolated for that *S*.

Role-Taking Conditions

Twenty-nine *Ss*, whose TAT productions and rankings met the above criteria, were called back

6 mo. later. On the opaque projector, they were shown a new stimulus card depicting an old woman and a young girl facing each other in an ambiguous relationship. All *Ss* were asked to make up two stories for this card. As the basis for each story, *S* was given one of her pairs of attribute statements. She was asked to assign each statement of the pair to whichever figure she wished and to construct a standard TAT story built around the figures so endowed. After both stories were completed, she was then required to take the roles of her characters following the standard RTT procedure.

The *S* was given two booklets which had been individually prepared for her. The two attribute statements were pasted onto the top of the first page of each booklet. Although all *Ss* were treated in the same manner during the testing, the booklets were numbered so that half of the *Ss* made up stories for the pair of defensively isolated attributes first, and the other half for the pair of nonisolated attributes first. For each pair of attribute statements the instructions were as follows:

In this test for creative imagination, you are again to make up a dramatic story about a picture which has two characters in it. However, this time there is an additional instruction. Attached to the first page of each booklet are two qualities or characteristics. In making up your story, you must assign one quality to each character.

Tell what led up to the event shown in the picture, describe what is happening at the moment, and tell how it will end. Include what each character is feeling and thinking. You will have about five minutes for each story.

After they had finished writing the initial stories, *Ss* were given the role-taking part of the task; that is, they were asked to go back and retell each story from the perspective of each of the two characters in the story. For the first story, the role of the young woman was assigned first; and for the second story, the role of the old woman was assigned first. Detailed instructions for group administration of the RTT are given in Feffer and Suchotliff (1966).

RTT decentering categories. As has been noted, role-taking performance is evaluated in terms of the degree to which *S* is able to refocus upon his initial story from the perspectives of his characters while at the same time maintaining continuity between these various versions of the initial story. The scoring categories⁴ of the RTT are ordered along a di-

⁴ A detailed description of the RTT procedure and scoring criteria is available in mimeographed form from the authors or for a fee from the American Documentation Institute. Order Document No. 9010 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$2.50 for microfilm or \$6.25 for photocopies and make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

SIGNIFICANCE TEST FOR DECENTERING UNDER THE DIFFERENT CONDITIONS

Mean RTT score ^a		N	s	p ^b
Defensive isolation condition	Nonisolation condition			
11.5	14.4	29	3.0	.001

^a Mean scores are presented for descriptive purposes and were not used as a basis for statistical inference. They were derived by assigning weights to the role-taking task (RTT) categories.

^b Probability estimate is one-tailed (Wilcoxon Matched Pairs Signed Ranks test).

mention of increasing subtlety of coordination between *S*'s versions of the initial story. Categories at one extreme reflect obvious inconsistency or discontinuity between the characters' viewpoints as, for example, *S*'s describing the woman as "angry" in the initial story and as "happy" when taking the woman's role (simple refocusing). At the other extreme, the highest category (change of perspective) reflects an overall synthesis of two different perspectives. For example, *S* in the role of one woman may describe herself from an internal orientation and the other from an external viewpoint: "She feels cheated because she thought that the landlady had agreed not to raise her rent, but now she went ahead and raised it." In taking the role of the other woman, the inner and outer orientations are reversed, but at the same time coordinated: "The landlady could see that she was angry, but she had no choice—expenses had gone up."

Role-taking performance was scored in terms of the highest category attained in each story situation. Interjudge reliability was determined by the degree of correlation between the scoring of the first and second authors. The Spearman rank difference correlation was +.69 ($N = 20$, $p < .01$). Further information relevant to interjudge and test reliability may be found in Feffer (1959), Feffer and Gourevitch (1960), and Feffer and Jahelka (1968).

RESULTS

As may be seen in Table 1, RTT scores are significantly lower under the defensive isolation condition than under the nonisolation condition. In terms of the measures used, then, *S*s evidenced a greater degree of sequential decentering when required to take the viewpoints of defensively isolated attributes than when required to take the viewpoints of nonisolated attributes.

Since the decentering measure as well as the judgments of defensive isolation are based upon stories written by *S*, the possibility was

considered that the findings were influenced by a productivity variable. For example, being required to deal in writing with self-attributes which are unacceptable and/or unacknowledged may result in the desire to escape from the situation by writing less, this in turn resulting in a lower level of RTT functioning. In order to examine this possibility, verbal productivity under the various conditions was assessed. No significant differences were found in the number of words produced under the different conditions.

DISCUSSION

The present findings can be interpreted as supporting the hypothesis that a failure to coordinate perspectives occurs when these perspectives represent defensively isolated aspects of the self-organization. This interpretation is based, in part, upon the acceptance of the projective assumption that any personality characteristic which the individual ascribes to unstructured material is an attribute of the self. Clearly, however, interpretations other than those based upon the projective assumption can be advanced as, for example, the possibility that *S*s might have had difficulty in role taking solely because they were required to assume viewpoints which they disliked. In addition, it should be noted that even if one were to accept the projective assumption, it does not follow that the present findings are interpretable only in terms of the anxiety-defense paradigm set forth in the rationale. One could argue with equal justification that forcibly confronting the individual with an anxiety-provoking attribute in role taking produces not a defensive reaction of isolation, but a disorganized, undifferentiated response which is directly manifested in a lack of coordination between perspectives.

These problems obviously require further study. Should such research, however, consolidate the position that defensive isolation is basic to a difficulty in "conserving" role relationships on the RTT, then this would focus attention on a question of considerable import; namely, what implication does this type of isolation have for the development of conservation in general, including the conservation of impersonal reality?

Within Piaget's framework, the conditions for the development of conservation have been articulated in terms of an equilibration model. Essentially, this model suggests that the individual seeks contact with the environment in order to consolidate modes of organizing experience. Given the inherent properties of an average environment, however, the individual encounters internal contradictions between the ways in which he has previously organized experience. This leads to a cognitive reorganization whereby the internal conflict is resolved; the cycle then repeats itself.

Diametrically opposed to the internal dialectic of Piaget's equilibration process are the assumptions of an external reinforcement, drive-reduction model which, as White (1959) has systematically noted, underlie the anxiety-defense paradigm. The contrast between the equilibration and drive-reduction models has been further underscored by the implication (Schachtel, 1954; White, 1959; Wolff, 1960) that they refer to essentially antagonistic modes of behavior. These writers have observed that it is precisely when strong drives are *not* in operation that the organism is directed toward the exploration of new facets of the environment. In the presence of strong motivation, including anxiety, exploratory behavior gives way to a narrow form of learning specifically directed toward reduction of the high drive state. Such behavior, in supplanting exploration, may thus isolate the individual from further contact with new, potentially contradictory, dimensions of experience necessary to further development.

These conceptions would imply that exploratory behavior is necessary for the development of conservation in both impersonal and interpersonal realms. It seems reasonable to assume, however, that the world of interpersonal relationships differs from the world of objects in having a greater potential for eliciting strong motive and anxiety states. Therefore, there may be a greater number of occasions in the interpersonal sphere for exploratory behavior to give way to a drive-reducing, isolating form of learning. From this viewpoint, defensive isolation can be seen as behavior, which in supplanting exploratory activity, prevents the development of conservation in particular role relationships.

The extension of the decentering concept to the interpersonal area provides a framework for evaluating this suggested relationship between isolation due to dynamic, defensive factors, and the development of conservation. For example, one could compare training procedures based on "cognitive conflict" to those based on external reinforcement with respect to their relative effectiveness in inducing conservation on both the RTT and impersonal tasks (cf. Gruen, 1965; Smedslund, 1961a, 1961b; Wohlwill & Lowe, 1962). In addition, levels of motivation and anxiety could be varied in order to evaluate whether the effectiveness of cognitive conflict training is more readily disrupted on the RTT than on impersonal tasks. Such studies in providing information regarding the suggested relationship between dynamic factors and the equilibration model would, by the same token, be relevant to Wolff's (1960) conceptualization of the relationship between the classical Freudian view of development and Piaget's "conflict-free" approach to the problem.

REFERENCES

- CANDELL, P. Social egocentrism as a function of cognitive development and social withdrawal. Unpublished honor's thesis, Clark University, 1965.
- FEFFER, M. The cognitive implications of role-taking behavior. *Journal of Personality*, 1959, 27, 152-168.
- FEFFER, M. Symptom expression as a form of primitive decentering. *Psychological Review*, 1967, 74, 16-28.
- FEFFER, M., & GOUREVITCH, V. Cognitive aspects of role-taking in children. *Journal of Personality*, 1960, 28, 383-396.
- FEFFER, M., & JAHIELKA, M. Implications of the decentering concept for the structuring of projective content. *Journal of Consulting and Clinical Psychology*, 1968, 32, 434-441.
- FEFFER, M., & SUCHOTLIFF, L. Decentering implications of social interaction. *Journal of Personality and Social Psychology*, 1966, 4, 415-423.
- GRUEN, G. E. Experiences affecting the development of number conservation in children. *Child Development*, 1965, 36, 963-979.
- LOWENHERZ, L. Cognitive level as a function of defensive isolation. Unpublished doctoral dissertation, Yeshiva University, 1966.
- PIAGET, J. *The psychology of intelligence*. New York: Harcourt Brace, 1950.
- PILL, E., MISCELL, W., & BERNSTEIN, L. A note on remoteness of TAT figures as an interpretive concept. *Journal of Consulting Psychology*, 1959, 23, 252-255.

- SCHACHTEL, E. G. The development of focal attention and the emergence of reality. *Psychiatry*, 1954, 17, 309-324.
- SMEDSLUND, J. The acquisition of conservation of substance and weight in children. II External reinforcement of conservation of weight and of the operations of addition and subtraction. *Scandinavian Journal of Psychology*, 1961, 2, 71-83. (a)
- SMEDSLUND, J. The acquisition of conservation of substance and weight in children. V. Practice in conflict situations without external reinforcement. *Scandinavian Journal of Psychology*, 1961, 2, 158-160. (b)
- WHITE, R. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, 66, 297-313.
- WOHLWILL, J. F., & LOWE, R. C. An experimental analysis of the development of the conservation of number. *Child Development*, 1962, 33, 153-167.
- WOLFE, R. The role of conceptual systems in cognitive functioning at varying levels of age and intelligence. *Journal of Personality*, 1963, 31, 108-124.
- WOLFE, P. The developmental psychologies of Jean Piaget and psychoanalysis. *Psychological Issues*, 1960, 2(1, Whole No. 5).

(Received December 31, 1967)

RESPONSES OF PHENYLKETONURIC CHILDREN ON A CONTINUOUS PERFORMANCE TEST¹

V. ELVING ANDERSON,² FELICIA S. SIEGEL, ROBERT O. FISCH, AND ROBERT D. WIRT

University of Minnesota

A task which required continuous attending was given to 11 phenylketonuric children and 11 nonphenylketonuric control Ss (matched for age, IQ, sex, and race). The phenylketonuric Ss made significantly fewer correct responses and significantly more late responses than their matched controls. The performance of the phenylketonuric Ss also changed significantly over the sequence of stimulus presentations, with fewer correct responses toward the end.

The evidence that has been collected to date indicates that phenylketonurics treated early (prior to 4 mo. of age) do not attain an intelligence level commensurate with that of their unaffected siblings (Berman, Waisman, & Graham, 1966). It has not yet been determined whether further modifications in dietary treatment would eliminate the discrepancy. A further question which is raised by the results of the Berman et al. study is whether deficits are present in specific areas of functioning and, if there are deficits, whether they are shared by both treated and untreated phenylketonurics.

The results of a study on delayed response in phenylketonurics (Siegel, Anderson, Fisch, & Brown, 1967)² suggested that phenylketonurics were more affected by increased delay on a color-form task than their matched controls. More errors were made by both treated and untreated phenylketonurics. The indication that there may be a specific deficit in attention span led to the present study with the purpose of using a different method to investigate attention in order to help clarify the extent of the deficit.

The method used in the present study was based on that of Rosvold, Mirsky, Sarason, Borsome, and Beck (1956) and Mirsky and Van Buren (1965). In order to detect lapses

in attention Rosvold et al. developed a task which required continuous attending. For the present study the task was modified since the original task required the ability to readily discriminate letters of the alphabet, which some of the Ss were not able to do.

METHOD

Subjects

The experimental sample of 11 phenylketonuric Ss was obtained through the Phenylketonuric Clinic at the University of Minnesota Hospitals (Table 1). All experimental Ss were considered to have classical phenylketonuria since they met the following criteria: (a) They all had elevated fasting serum phenylalanine levels (21-43 mg. %) while on a regular diet. (b) They had normal or low serum-tyrosine values on tests during the previous year. (c) When having an elevated serum-phenylalanine level, they excreted in the urine an increased amount of phenylalanine and its metabolites.

In addition, 7 of the 11 Ss had sibs with a diagnosis of phenylketonuria at the time of study. The Ss 1E, 12E, and 14E (Table 1) were albs. Ss 4E, 7E, 8E, 11E, and 13E were from different families, but had phenylketonuric sibs not included in the study. None of the other Ss had phenylketonuric relatives.

The ages at which a low-phenylalanine diet was initiated or discontinued are also shown in Table 1. The effectiveness of dietary control at the time of the study was rated according to the criteria used by Koch, Acosta, Flashier, Schaeffler, and Wohlers (1967).

A control S matched for age, sex, IQ, and race with each experimental S was obtained through the Minneapolis Public Schools (Table 1). An additional requirement was that control Ss have no obvious motor difficulty such as cerebral palsy or other physical impairment. Ss 1C, 2C, 6C, 8C, and 10C had IQs over 80. The medical diagnosis for S 12C was Down's syndrome. The medical records for the remaining control Ss did not provide information that would help to establish the etiology of their retardation.

¹ This investigation was supported by Research Grant HD 01396 from the National Institute of Child Health and Human Development.

² Requests for reprints should be sent to V. Elving Anderson who is now at the Dight Institute for Human Genetics, Zoology Building, Minneapolis, Minnesota 55455.

³ F. S. Siegel, V. E. Anderson, R. O. Fisch, & B. W. Brown. Unpublished manuscript, 1967.

TABLE 1
CHARACTERISTICS OF THE EXPERIMENTAL AND CONTROL SUBJECTS

Matched pair	Sex	Age at study		IQ		Special diet in experimental Ss		
		E	C	E	C	Age on	Age off	Current control
1E 1C	F	6-6	6-2	95	95	12 days		Control
2E 2C	M	7-0	6-11	92	94	0-3		Control
4E 4C	M	8-4	8-1	76	76	6-10	7-11	
6E 6C	M	8-7	8-7	104	107	0-9		Control
7E 7C	M	10-7	10-0	75	74	0-6	9-0	
8E 8C	M	10-8	10-9	85	85	0-2		Control
10E 10C	F	11-2	11-6	92	96	2-0		Excellent
11E 11C	F	11-10	12-4	61	69	1-4	5-0	
12E 12C	M	13-5	13-2	43	42	Never on diet		
13E 13C	M	13-9	13-7	78*	77*	11-6	12-9	
14E 14C	F	15-9	16-2	50	48	Never on diet		
M		10-8	10-7	77.4	78.4			

Note.—E = experimental Ss; C = control Ss.
* WISC; all others by Stanford-Binet.

Apparatus

Eight picture vocabulary items selected from the 1960 Stanford-Binet revision were made into 2 × 2 in. slides.⁴ The slides were projected by a Carousel projector onto a screen 9 × 6 in. and were automatically presented at 1.9-sec. intervals. The slide was exposed 1.2 sec. with a dark interval of .7 sec. The screen was mounted on the surface of a wooden enclosure 19 × 14 × 13 in. The slides were projected from the rear, thus hiding the projector from S's view.

Through the use of photoelectric cells and relays it was possible to record five types of responses: the correct response and four types of errors. The timer and the five counters were in a separate unit in view of E.

The S was seated 4 ft. away from the screen and E was seated behind S and out of his line of vision. The S held a plastic unit 18 × 3 × 2 in. on which the response button was mounted.

Procedure

All Ss were familiar with the picture items from a previous test. The first slide presented showed a tree, and the following instructions were given:

I am going to show you some pictures. One of the pictures is a tree. Every time you see a tree press the button. [E handed the switch to S and observed S to determine that the correct procedure was used.] Let's try a few. Remember, press the button every time you see a tree.

Following the instructions 10 slides were presented including 3 possible correct stimuli. When S responded correctly to all stimuli, testing was begun. A total of 120 slides, including 30 correct stimuli,

were presented. In this series S's responses were recorded as correct, incorrect, or late. A late response was one which occurred between slide presentations, that is, in the blank interval. Upon completion of the tree (T) task S was given a rest period and allowed to leave his chair.

When testing was resumed, the following instructions were presented. "Now we are going to do something a little bit different. Press the button when you see a tree but only if the tree comes right after a ball. If you see a ball and then a tree, press the button. Let's try a few." Eleven pretest slides were presented which included three correct combinations plus possible errors such as a tree not preceded by a ball or a ball not followed by a tree. Incorrect responses were discussed and the testing was not initiated until S performed correctly. Before starting testing S was told, "We're going to do some just like that. Press the button when you see a tree, if it comes right after a ball." On the ball-tree (B-T) task no S required more than three presentations of the pretest series of slides.

For the B-T task 201 slides were presented including 30 possible correct stimuli and 171 stimuli where a tree was not preceded by a ball. For the B-T task five counters recorded responses as follows: (a) correct, (b) tree not after ball, (c) ball, (d) other incorrect responses and (e) late responses. Responses were also recorded by E as they occurred in order to obtain information about sequence. Hand recording was not as reliable as the automatic counters. Exact information was obtained for the sequence of responses to the correct stimulus while there were some missing data for the sequence of incorrect responding.

Testing time for the T task was 14 min. B-T required 64 min. Total testing time including instructions, pretest, and rest time was approximately 20 min.

⁴ Permission to make these slides was granted by the Houghton-Mifflin Company.

RESPONSES OF PHENYLKETONURIC CHILDREN ON A CONTINUOUS PERFORMANCE TEST¹

V. ELVING ANDERSON,² FELICIA S. SIEGEL, ROBERT O. FISCH, AND ROBERT D. WIRT

University of Minnesota

A task which required continuous attending was given to 11 phenylketonuric children and 11 nonphenylketonuric control Ss (matched for age, IQ, sex, and race). The phenylketonuric Ss made significantly fewer correct responses and significantly more late responses than their matched controls. The performance of the phenylketonuric Ss also changed significantly over the sequence of stimulus presentations, with fewer correct responses toward the end.

The evidence that has been collected to date indicates that phenylketonurics treated early (prior to 4 mo. of age) do not attain an intelligence level commensurate with that of their unaffected siblings (Berman, Waisman, & Graham, 1966). It has not yet been determined whether further modifications in dietary treatment would eliminate the discrepancy. A further question which is raised by the results of the Berman et al. study is whether deficits are present in specific areas of functioning and, if there are deficits, whether they are shared by both treated and untreated phenylketonurics.

The results of a study on delayed response in phenylketonurics (Siegel, Anderson, Fisch, & Brown, 1967)² suggested that phenylketonurics were more affected by increased delay on a color-form task than their matched controls. More errors were made by both treated and untreated phenylketonurics. The indication that there may be a specific deficit in attention span led to the present study with the purpose of using a different method to investigate attention in order to help clarify the extent of the deficit.

The method used in the present study was based on that of Rosvold, Mirsky, Sarason, Bransome, and Beck (1956) and Mirsky and Van Buren (1965). In order to detect lapses

in attention Rosvold et al. developed a task which required continuous attending. For the present study the task was modified since the original task required the ability to readily discriminate letters of the alphabet, which some of the Ss were not able to do.

METHOD

Subjects

The experimental sample of 11 phenylketonuric Ss was obtained through the Phenylketonuric Clinic at the University of Minnesota Hospitals (Table 1). All experimental Ss were considered to have classical phenylketonuria since they met the following criteria: (a) They all had elevated fasting serum phenylalanine levels (21-43 mg. %) while on a regular diet. (b) They had normal or low serum-tyrosine values on tests during the previous year. (c) When having an elevated serum-phenylalanine level, they excreted in the urine an increased amount of phenylalanine and its metabolites.

In addition, 7 of the 11 Ss had sibs with a diagnosis of phenylketonuria at the time of study. The Ss 1E, 12E, and 14E (Table 1) were sibs. Ss 4E, 7E, 8E, 11E, and 13E were from different families, but had phenylketonuric sibs not included in the study. None of the other Ss had phenylketonuric relatives.

The ages at which a low-phenylalanine diet was initiated or discontinued are also shown in Table 1. The effectiveness of dietary control at the time of the study was rated according to the criteria used by Koch, Acosta, Fishler, Schaeffler, and Wohlers (1967).

A control S matched for age, sex, IQ, and race with each experimental S was obtained through the Minneapolis Public Schools (Table 1). An additional requirement was that control Ss have no obvious motor difficulty such as cerebral palsy or other physical impairment. Ss 1C, 2C, 6C, 8C, and 10C had IQs over 80. The medical diagnosis for S 12C was Down's syndrome. The medical records for the remaining control Ss did not provide information that would help to establish the etiology of their retardation.

¹ This investigation was supported by Research Grant HD 01396 from the National Institute of Child Health and Human Development.

² Requests for reprints should be sent to V. Elving Anderson who is now at the Dight Institute for Human Genetics, Zoology Building, Minneapolis, Minnesota 55455.

³ F. S. Siegel, V. E. Anderson, R. O. Fisch, & B. W. Brown. Unpublished manuscript, 1967.

TABLE 1
CHARACTERISTICS OF THE EXPERIMENTAL AND CONTROL SUBJECTS

Matched pair	Sex	Age at study		IQ		Special diet in experimental Ss		
		E	C	E	C	Age on	Age off	Current control
1E 1C	F	6-6	6-2	95	95	12 days		Good
2E 2C	M	7-0	6-11	92	94	0-3		Good
4E 4C	M	8-4	8-1	76	76	6-10	7-11	
6E 6C	M	8-7	8-7	104	107	0-9		Good
7E 7C	M	10-7	10-0	75	74	0-6	9-0	
8E 8C	M	10-8	10-9	85	85	0-2		Good
10E 10C	F	11-2	11-6	92	96	2-0		Excellent
11E 11C	F	11-10	12-4	61	69	1-4	5-0	
12E 12C	M	13-5	13-2	43	42	Never on diet		
13E 13C	M	13-9	13-7	78*	77*	11-6	12-9	
14E 14C	F	15-9	16-2	50	48	Never on diet		
M		10- 8	10-7	77.4	78.4			

Note.—E = experimental Ss; C = control Ss.
* WISC; all others by Stanford-Binet.

Apparatus

Eight picture vocabulary items selected from the 1960 Stanford-Binet revision were made into 2×2 in. slides.⁴ The slides were projected by a Carousel projector onto a screen 9×6 in. and were automatically presented at 1.9-sec. intervals. The slide was exposed 1.2 sec. with a dark interval of .7 sec. The screen was mounted on the surface of a wooden enclosure $19 \times 14 \times 13$ in. The slides were projected from the rear, thus hiding the projector from S's view.

Through the use of photoelectric cells and relays it was possible to record five types of responses: the correct response and four types of errors. The timer and the five counters were in a separate unit in view of E.

The S was seated 4 ft. away from the screen and E was seated behind S and out of his line of vision. The S held a plastic unit $1\frac{1}{2} \times 3 \times \frac{1}{2}$ in. on which the response button was mounted.

Procedure

All Ss were familiar with the picture items from a previous test. The first slide presented showed a tree, and the following instructions were given:

I am going to show you some pictures. One of the pictures is a tree. Every time you see a tree press the button. [E handed the switch to S and observed S to determine that the correct procedure was used.] Let's try a few. Remember, press the button every time you see a tree.

Following the instructions 10 slides were presented including 3 possible correct stimuli. When S responded correctly to all stimuli, testing was begun. A total of 120 slides, including 30 correct stimuli,

⁴ Permission to make these slides was granted by the Houghton-Mifflin Company.

were presented. In this series S's responses were recorded as correct, incorrect, or late. A late response was one which occurred between slide presentations, that is, in the blank interval. Upon completion of the tree (T) task S was given a rest period and allowed to leave his chair.

When testing was resumed, the following instructions were presented. "Now we are going to do something a little bit different. Press the button when you see a tree but only if the tree comes right after a ball. If you see a ball and then a tree, press the button. Let's try a few." Eleven pretest slides were presented which included three correct combinations plus possible errors such as a tree not preceded by a ball or a ball not followed by a tree. Incorrect responses were discussed and the testing was not initiated until S performed correctly. Before starting testing S was told, "We're going to do some just like that. Press the button when you see a tree, if it comes right after a ball." On the ball-tree (B-T) task no S required more than three presentations of the pretest series of slides.

For the B-T task 201 slides were presented including 30 possible correct stimuli and 30 stimuli where a tree was not preceded by a ball. For the B-T task five counters recorded responses as follows: (a) correct, (b) tree not after ball, (c) ball, (d) other incorrect responses, and (e) late responses. Responses were also recorded by E as they occurred in order to obtain information about sequence. Hand recording was not as reliable as the automatic counters. Exact information was obtained for the sequence of responses to the correct stimulus, while there were some missing data for the sequence of incorrect responding.

Testing time for the T task was 3.8 min. B-T required 6.4 min. Total testing time including instructions, pretest, and rest time was approximately 20 min.

TABLE 2

NUMBER AND TYPE OF RESPONSES TO THE "TREE" SERIES BY EXPERIMENTAL AND CONTROL SUBJECTS

Matched pair	Responded to tree		Chose wrong stimulus		Responded late	
	E	C	E	C	E	C
1E 1C	28	27	1	1	1	1
2E 2C	27	29	1	2	2	2
4E 4C	29	25	0	1	2	3
6E 6C	22	30	1	0	6	0
7E 7C	28	29	0	2	4	0
8E 8C	30	30	2	1	0	0
10E 10C	29	29	0	0	0	0
11E 11C	30	30	2	1	1	0
12E 12C	19	28	2	1	2	1
13E 13C	30	29	0	0	0	2
14E 14C	28	30	0	2	0	1
Total	300	316	9	11	18	10
M	27.3	28.7	.8	1.8	1.6	.9

Note.—E = experimental Ss, C = control Ss.

RESULTS

Nearly all Ss performed well on the T task (Table 2). Very few correct responses were omitted and few errors were made.

The data from one pair of Ss (Pair 12), the most retarded, had to be omitted from the results on the B-T task. It was clear from their performance that they had not main-

tained the instruction set and had become confused during the course of the task. The data for the remaining 10 Ss are shown in Table 3.

Since matched pairs of Ss were used, it was possible to compare the number of correct responses (excluding late responses) by computing the *t* value for the differences between pairs. The *t* value of the difference was 5.01, significant at the .001 level. Thus, it appeared that the phenylketonuric group (which included Ss who were retarded, nonretarded, on diet, and off diet) made significantly fewer correct responses than their matched controls (Table 3).

Table 3 shows that the frequency of errors, other than late responses, was approximately the same in both groups. The occurrence of late responses, however, was significantly more frequent in the phenylketonuric group ($t = 3.38, p < .01$).

A late response could follow either a correct or incorrect stimulus. From the record on hand-scored tally sheets, an attempt was made to determine where the late responses had occurred. The experimental group made 47 late responses, 27 after a correct stimulus, 7 after an incorrect stimulus, but the remaining 13 late responses could not be placed in sequence. When the late responses known to

TABLE 3

NUMBER AND TYPE OF RESPONSES TO THE "TREE AFTER BALL" SERIES BY EXPERIMENTAL AND CONTROL SUBJECTS

Matched pair	Responded to tree after ball		Chose wrong stimulus						Responded late	
			Tree	Ball	Other	Tree	Ball	Other		
	E	C	E	E	E	C	C	C	E	C
1E 1C	16	25	1	0	0	8	0	0	3	3
2E 2C	25	25	3	0	0	5	2	0	7	1
4E 4C	17	16	3	3	4	0	0	0	11	0
6E 6C	22	29	0	0	0	1	0	0	5	2
7E 7C	17	26	5	2	2	1	0	0	9	0
8E 8C	25	30	1	0	0	0	0	0	0	0
10E 10C	21	28	2	0	1	2	0	0	2	0
11E 11C	23	28	1	1	0	1	0	0	3	0
13E 13C	23	30	1	0	1	7	0	1	5	0
14E 14C	19	25	0	1	0	4	0	1	2	1
Total	208	262	17	7	8	29	2	2	47	7
M	20.8	26.2	1.7	.7	.8	2.9	.2	.2	4.7	.7

Note.—E = experimental Ss; C = control Ss.

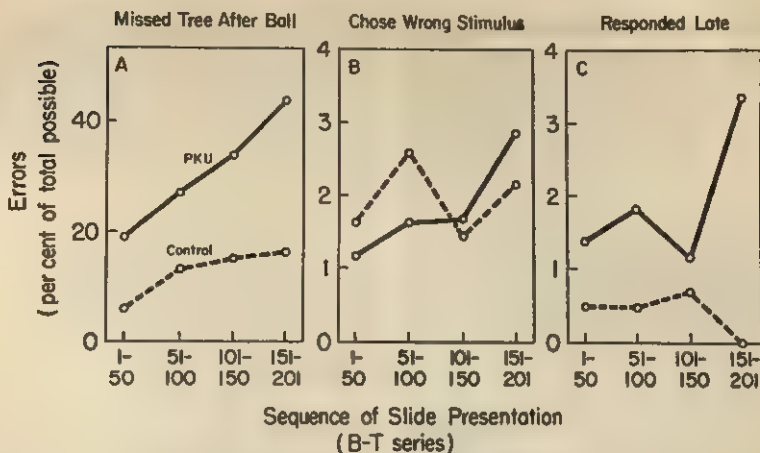


FIG. 1. Frequency of errors on the B-T series by type of error and by sequence of slide presentation.

follow the correct stimulus were added to the correct responses the difference between the two groups was eliminated.

Data showing the sequence of responses in blocks of 50 slide presentations are shown in Figure 1. The number of possible correct responses in each of the four blocks was 7, 7, 8, and 8, respectively.

Figure 1a shows that the experimental group made fewer correct responses as testing proceeded. The slope of the line for the experimental group was significantly different from zero ($t = 3.52$, $p < .01$). In the control group the t value was 1.61 ($p < .10$). All the experimental Ss had a positive slope whereas the control Ss tended to differ from each other.

Having found that the slope of correct responding over the stimulus presentations was different for the two groups, the question remained whether the two groups differed from the outset. Each experimental S's performance for the first 50 stimulus presentations was compared with that of his matched control. The t value of the difference between matched pairs was 3.86 which was significant at the .01 level.

Figure 1b presents the number of errors of commission made by each group and Figure 1c presents the late responses. Fourteen errors (1 error of commission and 13 late responses) made by the experimental group are missing from Figure 1. The errors were recorded by

the automatic counters, but were not located in sequence by hand scoring. There were no missing data for the control group. There appeared to be a sharp upswing in the number of errors, especially late responses, made by the experimental group toward the end of the test session. Other differences between the two groups might have emerged if the length of the test session had been extended.

In the analysis of a test designed to measure attention in children it would seem appropriate to expect a relationship between performance and mental age (MA). In the study (see Footnote 2) on delayed response a relationship between MA and performance was found. In the experimental group in the present study the correlation between MA and correct responses on B-T was $r = .41$ (ns). Analysis of the parallel data for the control group showed $r = .72$ ($p < .05$). The correlation between correct response on the B-T task and IQ was $r = .15$ (ns) for the experimental group and $r = .22$ (ns) for the controls. The relationship between CA and correct response was $r = .13$ for the experimental group and $r = .35$ for the controls.

DISCUSSION

In a task requiring continuous attending over a period of approximately $6\frac{1}{2}$ min. the performance of the phenylketonuric group in choice of correct stimulus was significantly below that of their matched controls. The

performance of the experimental group changed significantly over the sequence of stimulus presentations showing a highly significant positive slope of missed responses, whereas in the control group the slope was in the same direction but did not reach significance.

A comparison between the two groups on the initial level of performance also indicated that the experimental group made fewer correct responses from the beginning and from the slopes it appeared evident that the difference between the two groups increased as testing continued. Thus, it would appear that the phenylketonuric Ss not only showed an impaired attention span as compared to the control group, but they were also increasingly affected by the task which required constant attending. In an earlier study (see Footnote 2) in which discrete trials in delayed response were presented to the same Ss, only the performance of the experimental group declined over the course of the trials.

When late responses following a correct stimulus were added to the total correct responses, the performance of the two groups did not differ. The finding that most of the late responses occurred immediately after the correct stimuli suggested that experimental Ss, while maintaining a level of attention sufficient to recognize the correct sequence of stimuli, were not able to respond within the allotted time.

The errors and late responses which have been identified suggest that a sharp upswing in late responses in the experimental group occurred after 150 slide presentations. Thus one extension of the present study would be to increase the number of slide presentations in the B-T task.

The phenylketonuric Ss were a heterogeneous group consisting of some Ss with average intelligence, some retardates, some Ss on diet, and some off diet. Although the total sample was too small to evaluate the effect of intelligence level and diet on performance, it was clear from the results that the experimental Ss individually and as a group performed more poorly than their matched controls.

REFERENCES

- BERMAN, P. W., WAISMAN, H. A., & GRAHAM, F. K. Intelligence in treated phenylketonuric children: A developmental study. *Child Development*, 1966, 34, 731-747.
- KOCH, R., ACOSTA, P., FISHER, K., SCHAEFFLER, G., & WOHLERS, A. Clinical observations on phenylketonuria. *American Journal of Diseases of Children*, 1967, 113, 6-15.
- MIRSKY, A. F., & VAN BUREN, J. M. On the nature of the "absence" in centrencephalic epilepsy: A study of some behavioral, electroencephalographic and autonomic factors. *Electroencephalographic and Clinical Neurophysiology*, 1965, 18, 334-348.
- ROSVOLD, H. E., MIRSKY, A. F., SARASON, I., BRANSOME, E. D., JR., & BECK, L. H. A continuous performance test of brain damage. *Journal of Consulting Psychology*, 1956, 20, 343-351.

(Received December 28, 1967)

SENSORY-DECISION THEORY ANALYSIS OF THE PLACEBO EFFECT ON THE CRITERION FOR PAIN AND THERMAL SENSITIVITY (d')¹

W. CRAWFORD CLARK²

New York State Psychiatric Institute

Administration of a placebo described as a potent analgesic to 22 normal Ss sharply decreased the proportion of pain responses to radiant heat stimulation, suggesting that the threshold for pain had been raised. However, analysis of the data by sensory-decision theory demonstrated that thermal sensitivity (d') remained unchanged. The effect of the placebo was solely to alter Ss' response bias or likelihood ratio criterion ($p < .005$).

Quantification of the pain experience still eludes us. Even in the study of experimentally produced pain where the investigator has considerable control over a large number of variables such as stimulus parameters, the selection and training of Ss, and many other aspects of the experimental situation, there is very little agreement concerning either the threshold for pain or the effect of analgesics. Even one of the best procedures from the viewpoint of classical psychophysics, in which accurately determined amounts of radiant heat are applied to the skin according to the method of limits, yields widely varying results. Mean thresholds obtained by different investigators vary between 182 and 348 mcals/sec/cm² (Hardy, Wolff, & Goodell, 1952), and individual thresholds range from 170 to 472 mcals/sec/cm² (Clausen & King, 1950). Particularly distressing is the fact that the experimental pain threshold in man fails to respond dependably to powerful analgesic agents (Beecher, 1959). If the threshold is a valid indicator of sensory-physiological function, experimental pain thresholds would be expected to be raised by analgesics known to be effective relievers of pathological pain. Yet, even when the experimental pain experience has been made to resemble that of pathological pain by creating the slowly developing, sustained, aching pain produced by a tourniquet (Beecher, 1966), by ice-water

immersion (Wolff, Kantor, Jarvik, & Laska, 1966), tooth-pulp stimulation (Sherman, Fiasconaro, & Grundfest, 1963), or intramuscular injections of hypertonic saline (Jarvik & Wolff, 1962), the effects obtained with analgesics are inconsistent (Kutscher & Kutscher, 1957), and threshold variability remains a problem.

If physical and physiological variables cannot account for inconsistent responses to pain, perhaps psychological variables can. Non-sensory factors such as direct instructions, covert suggestion, social context, and even personality have been shown to exert a determining influence on the pain threshold (Hall, 1953). Instructions which induce a change in the criterion for pain alter the threshold for experimental pain (Wolff & Goodell, 1942); even a description of the sensations to be experienced produces a lower threshold (Clausen, Gjessvik, & Urdal, 1953). Suggestion, it is well known, will often reduce the anguish associated with pathological pain, particularly when augmented by a placebo (Hass, Fink, & Härtfelder, 1963). The importance of the experimental context has been demonstrated by McKenna (1958), who found that a large number of his Ss would not have regarded the noxious stimulus as painful had they encountered it outside the laboratory. Pain tolerance can also be influenced by social factors such as the saliency of group membership (Lambert, Libman, & Poser, 1960). With respect to personality variables, psychiatric patients whose predominant symptom was anxiety had a lower threshold for experimental pain stimulation

¹ This research was supported in part through Grant FR-05560 from the National Institute of Mental Health.

² Requests for reprints should be directed to W. Crawford Clark, New York State Psychiatric Institute, New York, New York.

TABLE 1
RATING SCALE FOR CATEGORIZING STRENGTH OF THERMAL EXPERIENCE
ALONG THE OBSERVATION CONTINUUM

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
R ₀ nothing	R ₁ maybe some- thing	R ₂ faint warmth	R ₃ warm	R ₄ moderate	R ₅ low heat	R ₆ hot	R ₇ very hot	R ₈ very faint pain	R ₉ faint pain	R ₁₀ pain	R ₁₁ very painful	R ₁₂ extremely painful (with- draw gun)

Note.—C and R designating criterion and response category were not in the scale viewed by S.

than did normal controls (Malmö & Shagass, 1949) or depressed patients (Hemphill, Hall, & Crookes, 1952). In addition, it has been reported that good tolerators of both pathological and experimental pain tended to be extraverted (Petrie, Collins, & Solomon, 1960) and to score high on a "psychopathic" index (Schalling & Kareby-Levander, 1963). Thus, threshold variability may be due not to changes in sensory sensitivity but to alterations in S's response bias—his general tendency to inhibit or to emit the response "pain." This possibility has been raised explicitly by Barber (1963), who concludes that although in some instances "hypnotically suggested analgesia" produces some amelioration of the pain experience, often the result is not a reduction in pain, but rather an unwillingness to admit to the *E* that pain had been experienced.

It should now be clear that pain thresholds obtained by traditional psychophysical procedures depend upon both sensory and non-sensory variables. Thus, when pain thresholds are reported to be functions of such variables as sex, fatigue, and anxiety level, in which physiological and psychological differences are intertwined, it is impossible to decide whether a difference in threshold represents altered sensory sensitivity or altered response bias. Fortunately, a new approach, based on the application of the theory of signal detection to human observers, has recently been developed (Green & Swets, 1966). The aspect of sensory-decision theory which is of particular importance here is the division of the traditional threshold into two components of observer performance. One component provides a relatively pure measure of sensory sensitivity (*d'*) which ideally remains un-

altered when nonsensory variables such as the subjective costs and values of yes and no are manipulated. The other independently determines the location of the observer's subjective criterion by relating it to the likelihood ratio (L_x). Using this procedure, Clark (1966) demonstrated that inhibiting instructions which induced a cautious attitude caused Ss to raise their criteria for visual flicker, while facilitating instructions had the opposite effect. In spite of this shift in response bias, sensory sensitivity (*d'*) remained unaltered, even though analysis of the same data by the method of constant stimuli indicated that the inhibiting instructions had raised the flicker threshold. In the present investigation, sensory-decision theory was used to study the effect of a placebo described as a potent analgesic on thermal sensitivity (*d'*) and the subjective criterion for pain.

METHOD

The 22 paid volunteers, mostly college students, each served in a control and a placebo session on different days. Sessions were counterbalanced to permit a statistical analysis of possible order effects. The Hardy-Wolff-Goodell Dolorimeter with its scale extended below 100 mc/sec/cm² to zero intensity was used to present radiant heat stimuli of 3-sec. duration to two 3.5-cm. diameter patches of India ink applied to either side of the midpoint between wrist and elbow on the volar surface of the non-preferred forearm. The approximate threshold of each S was determined by obtaining rating responses to stimuli of 200, 100, 260, 0, 300, and when necessary, 330 mc/sec/cm² in intensity, presented in the order listed. On the basis of these responses, a particular set of stimulus intensities tailored to each S was used throughout the experiment. The average values of the stimuli over all Ss were $S_0 = 0$, $S_1 = 94.4$, $S_2 = 208.6$, $S_3 = 271.0$, and $S_4 = 308.8$ mc/sec/cm². The stimuli, 25 at each intensity, were presented alternately to the distal and proximal spots every

TABLE 2
DRUG REACTION CHECKLIST

Item	Percentage of Ss giving a positive response		
	Start	During	End
1. Press left fingernail into the thumb; is there any loss of feeling?	27.3	50.0	27.3
2. Any feeling of an upset stomach or nausea?	4.5	13.6	0.0
3. Slight headache or heaviness in the head?	36.4	40.9	13.6
4. Any dizziness, or faintness?	13.6	9.1	13.6
5. Is your mouth drier than usual?	44.5	27.3	13.6
6. Do you feel high, light-headed, or euphoric?	18.2	31.8	9.1
7. Do you feel tense or nervous?	36.4	31.8	9.1
8. Do you notice any tingling sensations in the arm or elsewhere in the skin?	40.9	36.4	36.4
9. Press your teeth against your tongue. Does it feel numb?	22.7	40.9	18.2
10. Are you unable to concentrate?	4.5	18.2	13.6
11. Do you feel depersonalized, that is, that you're not really here, or you are outside yourself looking in?	22.7	13.6	9.1
12. Is there any feeling of bodily change or lightness?	27.3	18.2	9.1
13. Do you feel confused?	4.5	9.1	9.1

20 sec., the intensities being randomized with the restriction that each appear once every 15 trials. The following instructions were given:

We wish to determine your ability to feel warmth, heat, and faint pain. A variety of heat intensities, including zero, will be applied to your arm. Some stimuli will be so weak that you will feel nothing at all, others will be hot, while others will produce a very faint pain sensation. Remember, we don't want to see how much pain you can stand; rather, we want to know how good you are at detecting the presence of a just noticeable amount of pain. You may use as many descriptive categories as you wish. As an aid in maintaining consistency, here is a list of possible responses [see Table 1]. If the stimulus is too painful, you may remove the projector from the skin. Alternate the spots, starting with the one closer to the wrist.

During the placebo session, these additional instructions were given.

This drug is an analgesic or "pain-killer" which has been in use a long time and is, of course, completely safe. To date it has been used for internal, pathological pain. We are interested in determining its efficacy to reduce sensations of heat and pain produced by external thermal stimulation. I must first ask you a few questions to make certain that it is safe to administer the drug.

The questions on the medical checklist emphasized the importance of excellent health as a prerequisite to taking the drug, queried the strength of S's reaction to analgesics and anesthetics taken previously, and suggested that the drug was a narcotic of some sort by assuring him that a single dose could not be habit forming in any way. The S was asked if he were willing to sign a "release" at the end of the experiment, which included the statement that he was in complete control of himself and that he would not drive for the rest of the day. The S then swallowed a bitter, white powder consisting of 392-mg. lactose and 8-mg. quinine, followed by a glass of water.⁸ This amount of quinine is well below the minimum analgesic dose of 300 mg. (Goodman & Gilman, 1965). The idea that an analgesic had actually been administered was further strengthened by obtaining S's responses to a drug reaction checklist (see Table 2) near the beginning, middle, and end of the approximately 90-min. test session. The effectiveness of the induced mental set was demonstrated by the fact that all Ss but one reported some of the following effects: numbness of the tongue or fingers, slight headache, faint nausea, dizziness, depersonalization, confusion, and euphoria.

RESULTS

The group data for the cumulated probability of occurrence of any response (R) or a higher one (R+) conditional upon a particular stimulus intensity being present appear in Table 3. (Three Ss who failed to give any "nothing" responses had to be excluded from the study.) Isosensitivity curves may be generated from confidence-rating data by compressing the ratings to form a dichotomy at each of the confidence levels used to rate each of the stimuli (Hake & Rodman, 1966; Swets, Tanner, & Birdsall, 1961). In a similar manner—but substituting category ratings for confidence ratings—the responses in each of the rating categories, R_i ($i = 0$ to 12), to each of the five thermal intensities, S_0 to S_4 , were averaged over Ss and converted to conditional probabilities, that is, the probability of a particular response conditional upon a stimulus of a particular thermal intensity being present. Next, these conditional probabilities were cumulated successively from R_{12} to R_0 , about the criteria from C_{12} to C_1 to express the probability that a particular response or a higher one from the rating scale,

⁸The author thanks B. Montefusco, Senior Pharmacist at the New York State Psychiatric Institute, for providing the capsules and suggesting their composition and appearance.

TABLE 3
STIMULUS-RESPONSE MATRIX: CUMULATED PROBABILITY OF A RESPONSE
OR A HIGHER ONE TO A PARTICULAR STIMULUS

Stimulus intensity (mcal/sec/ cm ²)	Response category												
	R ₀₊ No+	R ₁₊ MS+	R ₂₊ FW+	R ₃₊ W+	R ₄₊ M+	R ₅₊ LH+	R ₆₊ H+	R ₇₊ VH+	R ₈₊ VFP+	R ₉₊ FP+	R ₁₀₊ P+	R ₁₁₊ VP+	R ₁₂ EP
A. Control condition													
S ₄ (308.8)		1.000	.998	.993	.982	.945	.905	.771	.716	.629	.445	.155	.044
S ₃ (271.0)		1.000	.994	.973	.938	.843	.738	.509	.444	.329	.162	.024	.002
S ₂ (208.6)	1.000	.993	.940	.766	.535	.358	.211	.087	.056	.011	.002		
S ₁ (94.4)	1.000	.760	.385	.114	.051	.025	.011		.004				
S ₀ (0.0)	1.000	.336	.064	.018	.007	.004	.002						
B. Placebo condition													
S ₄ (308.8)		1.000	.998	.974	.932	.898	.832	.658	.554	.414	.247	.102	.007
S ₃ (271.0)	1.000	.997	.975	.910	.837	.739	.608	.396	.307	.204	.098	.022	
S ₂ (208.6)	1.000	.984	.889	.720	.485	.320	.185	.075	.056	.020	.007		
S ₁ (94.4)	1.000	.609	.338	.095	.042	.024	.009		.007				
S ₀ (0.0)	1.000	.251	.044	.007	.005				.004				

Note.—Each row represents 550 responses, 25 to each stimulus by 22 Ss.

R₄₊ had occurred. It may be profitable to consider an example from Table 3, Section A. The cumulated conditional probability of the responses, R₁₊, that is, from "maybe something," "faintly warm," etc. to "extremely painful," locates the position of the rating criterion, C₁, used by the Ss to separate the sensory experience reflected in the response, "nothing" (R₀), from subjectively more intense sensory experiences. Thus for the zero-intensity stimulus, S₀, the cumulated probability is expressed as $p_{S_0}(R_1) = .336$. Similarly, the same criterion, C₁, applied to the stimulus of immediately higher intensity, S₁, yields the point $p_{S_1}(R_1) = .760$. These two values, which are analogous, respectively, to the false affirmative and hit rates obtained in the binary decision procedure, determine the upper right-hand point in quadrant S₁—S₀ of Figure 1; for this point, $d' = 1.13$, and $C_1 = .85$. C_x is analogous to the likelihood ratio criterion, L_x, and is determined in the same way—by dividing the probability that a particular observation value (x) arose from an S_i distribution ($i = 1$ to 4) by the probability that the same observation resulted from the adjacent lower distribution, S_{i-1}. In this manner, isosensitivity, or receiver-operating characteristic, curves (Figure 1) were con-

structed from the group data in Table 3, with response categories containing fewer than 10 responses being omitted, to determine whether the results follow the form predicted by sensory decision theory. The results are plotted on normal-normal coordinates, that is, on probability scales transformed so that the normal deviates are linearly spaced. The ordinate represents the cumulated probability that S perceived the more intense stimulus of a pair to be above a particular criterion value (analogous to the probability of a hit), while the abscissa represents the cumulated probability that he perceived the lower intensity stimulus to be above that same criterion (analogous to the probability of a false affirmative). For each of the four stimulus-intensity pairs (S₁—S₀, S₂—S₁, S₃—S₂, S₄—S₃) the points representing data obtained under both experimental conditions appear to fall on single lines with slopes close to unity. The underlying noise and signal-plus-noise distributions are therefore probably Gaussian, and of equal variance.

In order to make statistical tests to determine whether the placebo affected sensory sensitivity or response bias, separate stimulus-response matrices for control and placebo conditions were constructed for each S. From

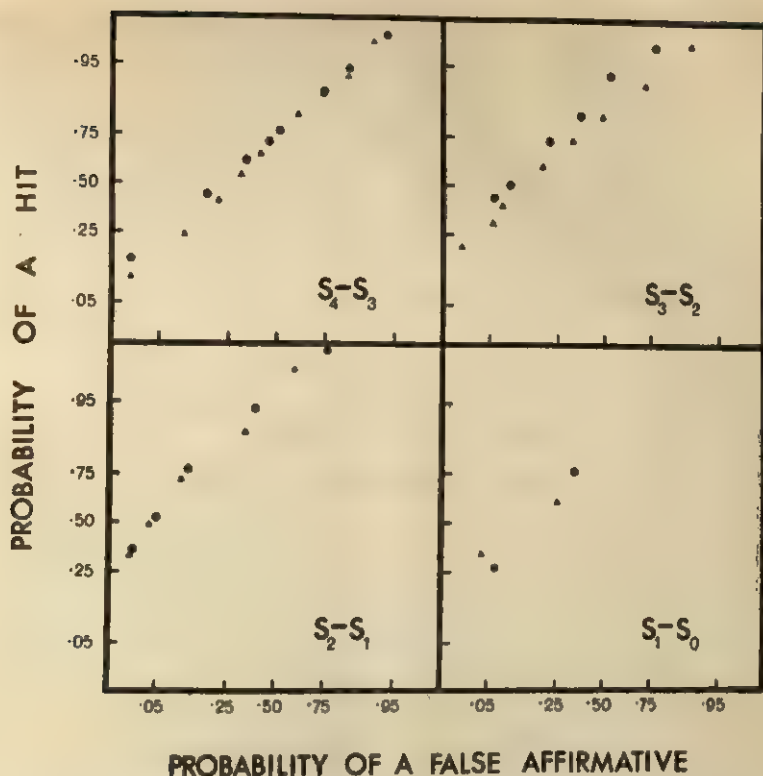


FIG. 1. Receiver-operating-characteristic curves under placebo (Δ) and control (O) conditions, plotted on normal-deviate axes. (The ordinate and abscissa obtained from the rating data are analogous to the hit and false-affirmative rates obtained from the binary decision technique.)

these, values of d' and C_x were computed for each S using the procedure just described for the group data. Since slightly different thermal intensities were used to stimulate each S , it was necessary to obtain a common measure of discriminability before pooling the data. This was done by taking the differences in intensity between each of the adjacent stimulus values (e.g., S_1-S_0 , S_2-S_1 , etc.) for each S and expressing them as mcal/sec/cm^2 per d' . The means of these values (somewhat akin to a just noticeable difference with response bias removed) for each of the four stimulus pairs under the placebo and control conditions appear in Table 4. The individual stimulus-response matrices also yielded the criterion measure (C_x) used by each S to judge his sensory experiences. The C_x is analogous to L_x , but reflects the fact that when more than two stimulus intensities are

being judged, criteria at different locations along the sensory continuum will have identical likelihood ratios (Luce, 1963). In this case, it must be assumed that S 's decision includes some additional factor, such as the magnitude of his sensory experience. Testing of the hypothesized shift in response bias

TABLE 4
MEAN THERMAL DISCRIMINABILITY

Condition	Mean discriminability (mcal/sec/cm^2 per d')			
	$(S_1-S_0)/d'$	$(S_2-S_1)/d'$	$(S_3-S_2)/d'$	$(S_4-S_3)/d'$
Control	83.25	65.33	40.60	31.55
Placebo	99.09	60.79	44.25	34.70
Mean over conditions	91.17	63.06	42.42	33.13

Note.— $S_0 = 0$, $S_1 = 94.4$, $S_2 = 208.6$, $S_3 = 271.0$, and $S_4 = 308.8$ mcal/sec/cm^2 .

TABLE 5
MEAN THERMAL CRITERIA (C_x) EXPRESSED
AS LIKELIHOOD RATIOS

Condition	Criterion			
	C_1 detection	C_2-C_4 warm	C_5-C_7 hot	C_8-C_{12} pain
Control	1.49	2.01	1.01	1.04
Placebo	1.90	2.34	1.75	1.76
Difference between conditions	0.41	0.33	0.74	0.72

between experimental conditions requires the measurement of the change in a particular criterion value to a particular stimulus pair, for example, C_{10} to S_4-S_3 . Thus only those criteria used by S on both sessions to the same stimulus pair were available for statistical analysis. However, no response data were lost, since values of C_x are cumulative measures. The Ss tended to associate particular responses, or criteria, with each of the stimulus intensities. For example, C_1 was frequently used when S_0 and S_1 were presented, and never resorted to when S_4 was presented. The "warm" criteria (C_2 , C_3 , and C_4) were used with S_2-S_1 , the "hot" criteria (C_5 , C_6 , and C_7) with S_3-S_2 , and the "pain" criteria (C_8 to C_{12}) were used to evaluate S_4-S_3 . The increase in mean criterion values induced by the placebo condition appears in Table 5.

According to Hartley's test there was no significant departure from homogeneity of variance, and an analysis of variance for repeated measures with Ss nested within order was used to treat separately the measures of discriminability, mcal/sec/cm^2 per d' , and of criterion magnitude, C_x (Winer, 1962). Thermal discriminability was not altered by the administration of the placebo, nor were there any significant order or interaction effects. However, discriminability did vary with stimulus intensity ($F = 23.3$, $df = 3/60$, $p < .001$), Ss becoming progressively more sensitive as intensity was increased. According to the Newman-Keuls procedure for the difference between treatment means, each of the mean values of mcal/sec/cm^2 per d' in the lower row of Table 4 differs significantly from each of the other values ($p < .01$) with the exception of 42.42 versus 33.13 mcal/sec/cm^2 per d' . Analysis of variance on C_x demon-

strated that administration of the placebo did significantly increase the magnitudes of the subjective criteria ($F = 8.53$, $df = 1/20$, $p < .01$). Each increment in the magnitude of Ss' mean criterion is significant ($p < .005$) according to t tests for correlated means (Table 5). Thus, not only were the criteria for "pain" (C_8 to C_{12}) raised, but the criteria for "hot," "warm," and even "detection" were also increased. There were no significant order or interaction effects for C_x .

A graphic portrayal of the results from Tables 4 and 5 appears in Figure 2. The abscissa is linear with respect to d' and may be regarded as the observation axis representing the magnitude of the observer's phenomenal experience (x), transformed to produce Gaussian density functions. The ordinate represents probability density. The left-hand distribution represents the probability that x will occur in the presence of no stimulus, and in the remaining distributions the probability that x will occur in the presence of one of the four thermal stimuli. Since the distributions overlap, each value of x has a set of probabilities, and hence of likelihood ratios, associated with it. The five density functions are assumed to be of equal variance, since there is no strong evidence that the slopes of Figure 1 depart significantly from unity. The fact that sensitivity to thermal stimulation (d') was not altered by the placebo is reflected in the single set of density functions averaged over experimental conditions. The placebo-induced increase in the value of each of the criteria (Table 5) is represented by the arrow. If a linear relationship between sensory experience and stimulus intensity over the range of 271.0 to 308.8 mcal/sec/cm^2 is assumed, then under the control condition the criterion for pain was set at the mean sensation magnitude produced by a stimulus of approximately 289 mcal/sec/cm^2 , while under the placebo condition the subjective criterion for pain was raised to 304 mcal/sec/cm^2 .

In order to compare the measures of d' and C_x with the thresholds obtained with traditional psychophysical procedures, the pain response data were also processed according to the method of constant stimuli. The psychophysical ogives were obtained by plotting the probability of a "faint-pain" (or greater)

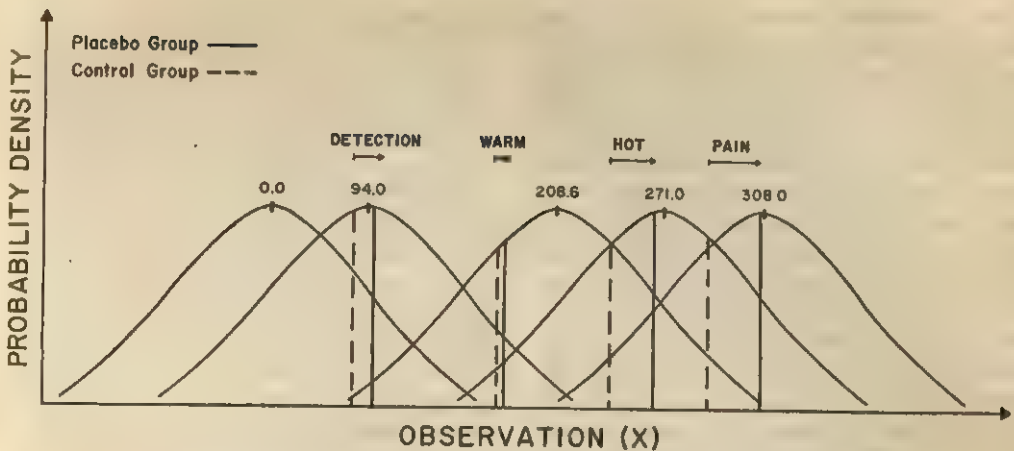


FIG. 2. Interlocking hypothetical probability density functions of S_0 , S_1 , S_2 , S_3 , and S_4 .

response against stimulus intensity, that is, $p_{S_2}(R_{S_+})$, $p_{S_3}(R_{S_+})$, and $p_{S_4}(R_{S_+})$ (see Table 3). The normally transformed data were readily fitted with a straight line. In contrast to the constancy of d' over experimental conditions, the constant stimuli procedure yielded interpolated 50% thresholds of 281 and 302 mcal/sec/cm² for the control and the placebo conditions, respectively.

DISCUSSION

The administration of a placebo described as a potent analgesic decreased the proportion of pain responses to intense thermal stimulation. (Although every S but one reported some effect from the capsule, there was no significant relation between the number of positive responses to the drug reaction checklist and the size of an individual's criterion shift according to Kendall's rank correlation coefficient, r .) Sensory-decision theory analysis indicates that the reduced number of pain responses reflects an increase in the amount of noxious stimulation S s were willing to endure before calling it pain, rather than a decrease in their thermal sensitivities. The constancy of the index of sensory sensitivity (d') strongly suggests that under the placebo condition the intensity of the phenomenal experience, or perhaps on a physiological level the number of neural impulses, produced by the stimulus did not decrease. The sole effect of the placebo was to cause S s to increase the magnitude of their criteria for pain (C_8

to C_{12}), presumably because the placebo-induced set increased the embarrassment or social cost of a "pain" response. In contrast to the constancy of d' , analysis of the same data by the method of constant stimuli led to the conclusion that the placebo raised the threshold to painful stimulation. Since the threshold is generally regarded as the reciprocal of sensory sensitivity, the customary interpretation would then be that the placebo had in fact exerted an analgesic effect and actually attenuated the experience of pain. According to decision theory analysis, however, the change in the constant stimuli threshold was due to the effect of suggestion on the attitude or response bias of S s. Other methods used to measure pain in which the subjective criterion of S is allowed to vary in an indeterminable manner, such as limits, adjustment, "titration," and response latency, are equally unsatisfactory. In sum, sensory-decision techniques offer a new approach to the study of pain perception, even though certain adaptations, such as the substitution of rating categories for a posteriori probability of stimulus presence, are necessary and may force the relaxation of some of the assumptions of the theory. The isosensitivity curves obtained demonstrate that S s were able to maintain at least an order relation between the 12 categories used, even when five different intensities were presented in a single session.

A number of researchers (Gelfand, 1964;

Wolff, Krasnegor, & Farr, 1965) who found that instructions altered pain tolerance but not pain thresholds have tentatively concluded that pain tolerance is determined primarily by psychological variables, while the pain threshold reflects the influence of purely sensory variables. The results of the present experiment, where the pain threshold was readily influenced by instruction-induced set, do not support this viewpoint. In fact, one may question whether a pain threshold, in the sense that threshold implies a lower limit on sensitivity, exists at all. The linearity and slope of the receiver-operating-characteristic curves obtained (Figure 1) support the sensory-decision concept that the criterion may be located anywhere along the continuous sensory-decision axis. Note, for example, that a few pain responses were obtained when the blank was presented.

Even when *E* does not intervene with a placebo, and in fact keeps his instructions as neutral as possible, different *S*s may perceive subtle interpersonal and social demand characteristics of the experimental situation in different ways (Orne, 1962; Rosenthal, 1963) and alter their pain criteria accordingly. Thus when the method of limits threshold for experimental pain is found to be higher in Northern Europeans than in Mediterranean people or Negroes, and higher in old than in young *S*s (Chapman & Jones, 1944), it is not clear whether these differences have a physiological or a psychological basis. Similarly, as Kolb (1962) has emphasized, differences in the reaction to pathological pain by patients who differ in age, socioeconomic status, race, or religion may be due to differences in attitude rather than differences in pain sensitivity.

Lower pain thresholds have been reported frequently in *S*s made anxious by the threat of shock, an impersonal *E*, unfamiliar apparatus, or inability to control the intensity of the pain stimulus (Haslam, 1966; Hill, Kornetsky, Flanary, & Wikler, 1952; Kornetsky, 1954; Lepanto, Moroney, & Zenhausern, 1965). Although it is possible that *S*s were experiencing more pain, a more reasonable explanation is that they had altered their response bias. The *S* who is worried that he might be hurt may say pain more often at

the lower intensities, perhaps in the interest of inducing *E* to avoid high-intensity stimuli, thus spuriously lowering his pain threshold. Obviously, it is impossible to control for all of the exquisitely subtle variables which may influence attitudinal bias. Even double-blind procedures fail to control for attitudinal shifts produced by the side effects of analgesics, such as nausea or euphoria. The best solution to this and other problems in pain perception appears to be to measure separately the sensory sensitivity and the attitudinal bias of each *S* by means of a sensory-decision theory procedure such as the one described here, or some variant of it.

REFERENCES

- BARBER, T. X. The effects of hypnosis on pain: A critical review of experimental and clinical findings. *Psychosomatic Medicine*, 1963, 25, 303-333.
- BEECHER, H. K. *Measurement of subjective responses: Quantitative effects of drugs*. New York: Oxford University Press, 1959.
- BEECHER, H. K. Pain: One mystery solved. *Science*, 1966, 151, 840-841.
- CHAPMAN, W. P., & JONES, C. M. Variations in cutaneous and visceral pain sensitivity in normal subjects. *Journal of Clinical Investigation*, 1944, 23, 81-91.
- CLARK, W. C. The psyche in psychophysics: A sensory-decision theory analysis of the effect of instructions on flicker sensitivity and response bias. *Psychological Bulletin*, 1966, 65, 358-366.
- CLAUSEN, J., GJESVIK, A., & URDAL, A. Changes in galvanic skin resistance as indication of pain threshold. *Journal of General Psychology*, 1953, 49, 261-271.
- CLAUSEN, J., & KING, H. E. Determination of the pain threshold on untrained subjects. *Journal of Psychology*, 1950, 30, 299-306.
- GELFAND, S. The relationship of experimental pain tolerance to pain threshold. *Canadian Journal of Psychology*, 1964, 18, 36-42.
- GOODMAN, L. S., & GILMAN, A. *The pharmacological basis of therapeutics*. New York: Macmillan, 1965.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HAKE, H. W., & RODWAN, A. S. Perception and recognition. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology*. New York: McGraw-Hill, 1966.
- HALL, K. R. L. Studies of cutaneous pain: A survey of research since 1940. *British Journal of Psychology*, 1953, 44, 279-194.
- HARDY, D. J., WOLFF, H. G., & GOODELL, H. *Pain sensations and reactions*. Baltimore: Williams & Wilkins, 1952.
- HASLAM, D. R. The effect of threatened shock upon

- pain threshold. *Psychonomic Science*, 1966, 6, 309-310.
- HASS, H., FINK, H., & HÄRTFELDER, G. Das placebo-problem. *Fortschritte der Arzneimittelforschung*, 1959, 1, 279-454. (Translation of selected parts in *Psychopharmacology Service Center Bulletin*, 1963, 2, 1-65.)
- HEMPHILL, R. E., HALL, K. R., & CROOKES, T. G. Preliminary report on fatigue and pain tolerance in repressive and psychoneurotic patients. *Journal of Mental Science*, 1952, 98, 433-440.
- HILL, H. E., KORNETSKY, C. H., FLANARY, H. G., & WIKLER, A. Effects of anxiety and morphine on discrimination of intensities of painful stimuli. *Journal of Clinical Investigation*, 1952, 31, 473-480.
- JARVIK, M. D., & WOLFF, B. B. Differences between deep pain responses to hypertonic and hypotonic saline solutions. *Journal of Applied Physiology*, 1962, 17, 841-843.
- KOLB, L. C. Symbolic significance of the complaint of pain. In W. H. Mosberg (Ed.), *Clinical neurosurgery*. Vol. 8. Baltimore: Williams & Wilkins, 1962.
- KORNETSKY, C. Effects of anxiety and morphine on the anticipation and perception of painful radiant thermal stimuli. *Journal of Comparative and Physiological Psychology*, 1954, 47, 130-132.
- KUTSCHER, A. H., & KUTSCHER, H. W. Evaluation of the Hardy-Wolff-Goodell pain threshold apparatus and technique: Review of the literature. *International Record of Medicine*, 1957, 170, 202-212.
- LAMBERT, W. E., LIBMAN, E., & POSER, E. G. The effects of increased salience of a membership group on pain tolerance. *Journal of Personality*, 1960, 38, 350-357.
- LEPANTO, R., MORONEY, W., & ZENHAUSERN, R. The contribution of anxiety to the laboratory investigation of pain. *Psychonomic Science*, 1965, 3, 475.
- LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosomatic Medicine*, 1949, 11, 9-24.
- McKENNA, A. E. The experimental approach to pain. *Journal of Applied Physiology*, 1958, 13, 449-456.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- PETRIE, A., COLLINS, W., & SOLOMON, P. The tolerance for pain and sensory deprivation. *American Journal of Psychology*, 1960, 73, 80-90.
- ROSENTHAL, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist*, 1963, 51, 268-283.
- SCHALLING, D., & KAREBY-LEVANDER, S. Ratings of anxiety-proneness and responses to electrical pain stimulation. Reports from Psychological Laboratory, No. 150, 1963, University of Stockholm.
- SHERMAN, H., FIASCONARO, J. E., & GRUNDFEST, H. Laboratory evaluation of analgetic effectiveness in human subjects. *Experimental Neurology*, 1963, 7, 435-456.
- SWETS, J. S., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, 68, 301-340.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WOLFF, H. G., & GOODELL, H. The relation of attitude and suggestion to the perception of and reaction to pain. *Proceedings of the Association for Research in Nervous and Mental Diseases*, 1942, 23, 434-448.
- WOLFF, B. B., KANTOR, T. G., JARVIK, M. E., & LASKA, E. Response of experimental pain to analgesic drugs. I. Morphine, aspirin, and placebo. *Clinical Pharmacology and Therapeutics*, 1966, 7, 224-238.
- WOLFF, B. B., KRASNEGOR, N. A., & FARR, R. S. The effect of suggestion upon experimental pain response parameters. *Perceptual and Motor Skills*, 1965, 21, 675-683.

(Received January 3, 1968)

VERBAL CONDITIONING OF COMMON ASSOCIATIONS IN LONG-TERM SCHIZOPHRENICS:

A FAILURE

A. STEVEN FRANKEL¹

AND

ALEXANDER M. BUCHWALD

University of Southern California

Indiana University

Ullmann, Krasner, and Edinger have stated that "common associations by schizophrenics to word association stimuli may be brought under control by selective reinforcement." The present study was an attempt to replicate and extend the Ullmann et al. study. The results suggested that social reinforcement of common associations by chronic schizophrenics fails to produce any evidence of a generalized increase in the tendency to give common associations. The results of the Ullmann et al. study may be seen as the result of a statistical artifact—regression to a common mean. Further, inappropriate use of the analysis of covariance (ANCOVA) to treat data of this type was discussed.

Ullmann, Krasner, and Edinger (1964) have stated that "common associations by schizophrenics to word association stimuli may be brought under control by selective reinforcement [p. 17]." They claim to have replicated and extended an earlier study by Sommer, Witney, and Osmond (1962). The latter presented a list of stimulus words to schizophrenics for a number of trials, reinforcing common associations with cigarettes and praise. They found that reinforcement led to an increase in the number of common associations given to *these* stimuli.

Ullmann and his co-workers administered the Kent-Rosanoff word list to two groups of 20 chronic schizophrenics. In order to obtain a base-line rate of common associations, neither group was reinforced for the first 20 responses. For the remaining stimuli, the experimental group was reinforced for each common association by a smile, "mmh-hmm," and a head nod. The control group was not reinforced for any response.

For the experimental group, which had been reinforced for common associations, Ullmann and his co-workers reported an increase in the number of common associations from the first 20 words on the Kent-Rosanoff list to the last 20 words. Such an increase did not occur for the nonreinforced controls, for whom

the number of common associations decreased from the first 20 words to the last 20.

Although Ullmann et al. claim to have replicated the results of the Sommer study, the two sets of data are really discrepant. Sommer and his co-workers tested their Ss for generalization on *new* stimulus words and were unable to find evidence that the learning effects had generalized. In contrast, in the Ullmann et al. study, each stimulus word was presented only once, and the effect of reinforcement was demonstrated on new stimulus words.

A closer look at the data reported by Ullmann et al. shows that the greatest mean differences between the two groups occurred on the first 20 trials. On this base-line measure, the control group gave more common association than the experimental group. This suggests that the results may be explained in terms of a statistical artifact—regression to a common mean. In other words, the operant level of the control group may have been too high due to random sampling, while the operant level of the control group was too low, due to random sampling. If this were the case, one would expect to find a decrease in the number of common associations given by the control group and an increase in the number of common associations given by the experimental group *even if reinforcement had no effect*. Further, the use of the analysis of covariance rests on the assumption that

¹ Requests for reprints should be sent to A. Steven Frankel, Psychology Department, University of Southern California, Los Angeles, California 90007.

treatment effects are uncorrelated with the covariate. A detailed discussion of this point is presented by Evans and Anastasio (1968) who note that violation of this assumption is serious when differences among the groups on the covariate are relatively large, particularly when only two treatment groups are involved. This was precisely the situation in the Ullmann et al. study. Since the base-line data were used in the analysis of covariance (ANCOVA) to adjust the final scores, the significant difference reported for these adjusted scores may have been spurious. To test this possibility, it was decided to replicate the Ullmann, Krasner, and Edinger experiment. An added control group was used in order to control for the effects of the occurrence of the reinforcement events independently of Ss' responses.

METHOD

Subjects. Sixty males, diagnosed schizophrenic without brain pathology, were assigned to one of three groups. These groups averaged 45 yr. of age, 10.5 yr. of education, and 15.8 yr. of hospitalization. The groups did not differ from each other significantly on any of these variables.

Materials. As in the Ullmann study, 98 words from the Kent-Rosanoff word list were used, with the Russell-Jenkins (1954) norms being used as a measure of common associations. Common associations were those which had been given by at least 100 Ss in the norm group.

Procedure. All Ss were administered the first 20 words of the Kent-Rosanoff word list without reinforcement, so as to establish a base line of common associations. For the remaining items, Group DRF (differential reinforcement) was reinforced for all common associations by a smile, "mmh-hmm," and a head nod. Group NRF (no reinforcement) was not reinforced for any responses, and Group CRF (continuous reinforcement) was reinforced for all responses, regardless of whether they were common or not. Except for the addition of the CRF group, the procedure employed in the present study was essentially similar to that used by Ullmann and his co-workers. Similarly, the measure of the effects of differential reinforcement was the number of common associations given to the last 20 words, adjusted for the number of common associations given to the first 20 items.

RESULTS

The analysis of variance performed on the number of common associations to the first 20 items by the three groups showed no significant differences in base-line rates. An analysis

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF COMMON ASSOCIATIONS GIVEN BY THREE GROUPS OF SCHIZOPHRENICS TO THE FIRST AND LAST 20 ITEMS OF THE KENT-ROSANOFF WORD LIST

Group	First 20 items		Last 20 items	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Differential reinforcement	10.7	2.83	10.01	2.99
Continuous reinforcement	7.65	5.15	7.2	4.7
No reinforcement	9.25	4.36	9.55	4.35

$$F = .96.$$

of covariance was performed on the number of common associations given to the last 20 stimulus words, using the base-line data as the covariance adjustment variable. The resulting F of .96 was not significant. Further, as may be seen in Table 1, the means do *not* seem to be in a direction which would suggest that conditioning took place.

DISCUSSION

These data do not support the Ullmann et al. (1964) hypothesis that "common associations by schizophrenics to word association stimuli may be brought under control by selective reinforcement [p. 17]." As suggested above, possible explanations for the difference between the results of the present study and those presented by Ullmann et al. lie in the inappropriate use of the analysis of covariance to treat the data, and in the possibility of a chance sampling error in the latter study. Given a spuriously high base line for the control group, it might be expected that a second measure, the responses to the last 20 stimuli, would show a decrease in the number of common associations. Similarly, given a spuriously low base line for the experimental group, it might be expected that a second measurement would show an increase in common associations.

What is there to suggest that reinforcement of common associations to certain words should increase the frequency of common associations to other words? This may be seen as an analogy to the case of bar pressing, in which reinforcement of a few instances brings about an increase in the rate of the response in general. This is true despite topographic

variations across individual response instances. Thus, these varying responses may be seen to be, or to have become, members of the same response class. But what is there to suggest that common associations constitute a single response class for long-term schizophrenics, or that they may become such a response class when reinforcement is given for specific responses? The present results and those of Sommer, Witney, and Osmond clearly do not. Other data seem equivocal. In general, the evidence indicates that attempts to teach schizophrenics to produce normal verbal behavior have resulted in limited amounts of generalization (e.g., Isaacs, Thomas, & Goldiamond, 1965; Sherman, 1965).

To summarize the findings of the present study, an attempt to replicate the study by Ullmann, Krasner, and Edinger was unsuccessful. Social reinforcement of common associations by chronic schizophrenics failed to produce any evidence of a generalized increase in the tendency to give common associations. It is suggested that the results obtained by Ullmann and his co-workers may have been

due to sampling vagaries rather than to experimental manipulations.

REFERENCES

- EVANS, S. H., & ANASTASIO, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin*, 1968, 69, 225-234.
- ISAACS, W., THOMAS, J., & GOLDIAMOND, I. Application of operant conditioning to reinstate verbal behavior in psychotics. In L. Ullmann & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test*. (DNR Tech. Rep. No. 11, Contract No. N8 ONR66216) Washington, D. C.: Government Printing Office, 1954.
- SHERMAN, J. Use of reinforcement and imitation to reinstate verbal behavior in mute psychotics. *Journal of Abnormal Psychology*, 1965, 70, 155-165.
- SOMMER, R., WITNEY, G., & OSMOND, H. Teaching common associations to schizophrenics. *Journal of Abnormal and Social Psychology*, 1962, 65, 58-61.
- ULLMANN, L., KRASNER, L., & EDINGER, R. Verbal conditioning of common associations in long-term schizophrenic patients. *Behavior Research and Therapy*, 1964, 2, 15-18.

(Received January 3, 1968)

SENSORY DEPRIVATION AND THE ENHANCEMENT OF HYPNOTIC SUSCEPTIBILITY

RAYMOND S. SANDERS, JR.¹ AND JOSEPH REYHER

Michigan State University

The purpose of the present study was to investigate the effect of sensory deprivation upon hypnotic susceptibility. Ten Ss initially resistant to hypnosis were placed in a light- and sound-attenuated sensory deprivation cubicle for a maximum of 6 hr. or until sensory deprivation phenomena were elicited. A hypnotic induction was undertaken via a communication system, while S remained in the deprivation cubicle, at a point when E judged that the sensory deprivation was sufficiently intense. Approximately 1 wk. following the experimental session, Ss submitted once again to a standard induction procedure. A control group of resistant Ss was included. Only the experimental group manifested a significant enhancement of hypnotic susceptibility, and this new level of susceptibility was maintained in the follow-up session.

Despite numerous investigations (Hilgard, 1965) into hypnotic susceptibility, understanding of individual differences in susceptibility is still limited to speculation and clinical inference. One possible method for identifying the critical variables is through the alteration of susceptibility.

Wiseman and Reyher (1962) reported that hypnotically induced dreams in conjunction with "deepening" suggestions increased the percentage of complete amnesia from 33 to 70 in a group of 30 Ss. However, this enhancement in performance on one task does not necessarily imply that performance would have been proportionately higher on other tasks. As, Hilgard, and Weitzenhoffer (1963) directly attacked the problem of modifying susceptibility by using a variety of deepening techniques frequently used by professionals in the field. They even spent time in psychotherapeutic attitude discussions but found that their efforts were rewarded by only slight gains as revealed by alternate forms of the Stanford Hypnotic Susceptibility Scale (SHSS). London (1967) concluded, "In general, Scale results are so consistent from session to session as to suggest that susceptibility is a very stable personality characteristic extremely resistant to change [p. 72]."

Several investigators have explicitly pointed to parallels between hypnotic induction and

sensory deprivation or perceptual isolation. Gill and Brenman (1959) suggested that all standard techniques of hypnotic induction involve extensive restrictions of meaningful sensory input in the visual, auditory, and kinesthetic modalities. These investigators concluded that conditions inherent in the standard hypnotic induction parallel the conditions imposed on Ss in the experiments with perceptual isolation initially reported by Bexton, Heron, and Scott (1954). Because of these apparent similarities between perceptual isolation and hypnotic induction, Gill and Brenman (1959) suggested that it would be valuable to study the effects of perceptual isolation on the hypnotizability of Ss. Also, Peña (1962) noted that the methods of achieving isolation or deprivation in an experimental setting are strikingly similar to the techniques used during the standard induction of hypnosis and that responses of isolated Ss resemble the behavior observed in hypnotic Ss, with evidence of heightened suggestibility being common to both situations.

Levitt, Brady, Ottinger, and Hinesley (1962) investigated the effect of experimentally induced sensory deprivation or perceptual isolation upon hypnotic susceptibility, but found no enhancement of susceptibility as a result of 4 hr. of sensory deprivation. Their Ss consisted of three female nursing students, refractory to hypnosis under standard conditions of induction, who, following the 4 hr. of deprivation, showed no enhancement of sus-

¹ Requests for reprints should be sent to Raymond S. Sanders, Student Health Service, Oregon State University, Corvallis, Oregon 97331.

ceptibility. The reactions of two of the *Ss* even seemed to be antagonistic to the relaxation and concentration required during hypnotic induction.

After 1½ and 3 hr. of perceptual isolation, however, Peña (1962) produced modest but significant enhancement in the susceptibility of two groups of prison inmates. He concluded that the effect was associated not only with perceptual isolation but also with the practice, implied suggestions, and cues incorporated in the procedure.

From his psychophysiological model of intrapsychic processes, Reyher (1964) reasons that hypnosis, sleep, sensory deprivation phenomena, and certain types of psychopathology are manifestations of the ascendance of lower levels of neural integration in the organization of brain functions and the regulation of behavior. Adaptive, integrated behavior is seen as a function of the dominance of the highest level of neuronal integration in the organization of brain functions. Any condition which eliminates or homogenizes sensory input prevents adaptive behavior, and the level of neuronal integration mediating adaptive behavior is replaced by a phylogenetically older and lower level of neuronal integration which becomes dominant in the overall organization of brain functions, producing an alteration of awareness. Intense anxiety, as a response to a hypnotic induction, or a defense against such anxiety engages adaptive behavior and prevents this reorganization from taking place. Reyher (1964, p. 117) also noted that sensory deprivation brings about a similar reorganization of brain mechanisms.

This psychophysiological model suggests that *S* should not be removed from sensory deprivation prior to a hypnotic induction, as removal reactivates adaptive behavior and its supporting level of neuronal integration. Rather, sensory deprivation should continue until *S* manifests disturbed ego functioning, which indicates, in this frame of reference, a putative change in the organization of brain mechanisms. Only resistant *Ss* were used in the present experiment, and the design included a follow-up hypnotic induction to assess the stability of any change in hypnotizability.

METHOD

Subjects

There were 10 volunteer college-age females in the experimental group who scored below 4 points on the Stanford Hypnotic Susceptibility Scale (SHSS), Form A. These resistant *Ss* were asked to return for two subsequent sessions and were told in advance that the second session involved 4-6 hr. of sensory deprivation after which they would again be hypnotized.

Materials and Experimental Setting

During the experimental session, each *S* was placed in a 12 × 12 ft. light- and sound-attenuated sensory deprivation cubicle. While reclining on a bed within the cubicle, *S* wore a headset consisting of a microphone and earphones which were part of an integrated communication system.² The remaining components of the system included a voice-activated tape recorder and an amplifier. The *Ss* were connected to a Grass No. 5 Polygraph (Yellow Springs Equipment Company) via electrodes secured to the first and second fingers of their left hands, which provided *E* with a continuous record of GSR activity and basal skin resistance. Pinned to the bed on which *S* reclined was a switch which activated a light in the adjacent equipment room. An auto-kinetic light served as a fixation object for *S* during the hypnotic induction which was live rather than recorded.

Instructions

Prior to *S's* entry into the deprivation cubicle, she was told that the session was designed to test the effect of sensory deprivation on her ability to be hypnotized. She was specifically instructed as follows:

Please tell me whatever you can about your reactions to this session. Your experiences while in sensory deprivation and subsequently those you have while being hypnotized will be helpful in understanding the results of the study. Whatever you choose to talk about will, of course, be kept strictly confidential. Also, try not to sleep during the session, but if you should doze off, don't feel guilty about it. At no time will I respond to what you say or answer any questions you may ask, or communicate with you in any way, unless you depress the switch on the bed. If you depress

² The microphone and earphones were manufactured by Midland Electronics under the trade name of Dynamic Headsets, model 21-236. The tape recorder was a Uher four-speed Report-L Recorder equipped with a voice-activating F411 Akustomat. Monitor and amplifier were built by Main Electronics under model No. SA1400, which included balance and volume controls to adjust recording levels. The integrated system was developed jointly by the Department of Psychology at Michigan State University and Main Electronics of Lansing, Michigan.

the switch, I shall read three stock market quotations to you. However, I shall always be in the adjacent room listening to what you say. I shall begin the hypnotic induction over the earphones at some point during the next 6 hours. After you are hypnotized, I shall enter the room to continue with some tasks similar to those we did the last time. Do you have any questions?

Procedure

While *S* was in the deprivation cubicle, *E* remained in the adjoining equipment room monitoring *S*'s spontaneous comments and reactions with the tape recorder and the polygraph. The *E* turned on the autokinetic light and began a hypnotic induction when *S* started to show sensory deprivation phenomena according to criteria developed from previous research (Sanders, 1967) with susceptible *S*s. These criteria were:

1. *Craving for stimulation* as indicated by (a) sharp increase in frequency of pressing button to hear stock market quotations, (b) statements denoting the wish for something to do or something to see, (c) self-stimulation, including singing, counting, tactile stimulation, etc., (d) comments on bodily sensations and alterations in body image, (e) asking *E* direct questions or demanding *E* respond to *S*, (f) statements of wanting to be hypnotized or to hear *E*'s voice as a signal the experiment is over, or wanting to quit, and (g) boredom and comments about the slow passage of time.

2. *Emotional lability* as indicated by (a) evidence of restlessness, fear, anxiety, or anger, (b) crying, weeping, or statements of depression, (c) marked fluctuation in mood (e.g., change from anxiety to feelings of well-being), (d) somatic disturbance, and (e) feelings of isolation and impending doom.

3. *Impaired secondary process and reality testing* as indicated by (a) marked impairment of the logical content in thought and speech, (b) hypnogogic imagery characterized as waking dreams, (c) hallucinatory experiences, somatic delusions, bodily disorientation, (d) personalistic interpretations, (e) increase in fantasy involvement, and (f) blank periods indicating inability to concentrate.

4. *Intensification of the relationship between E and S* as indicated by (a) statement of *S*'s feelings about *E*, (b) statements of personal problems and conflicts to *E*, (c) personal questions from *S* to *E*, (d) speculation about *E*, and (e) statements indicating need for contact with *E*.

The timing of the induction was the key point in the procedure and was largely a matter of clinical judgment based upon the above criteria. Any doubt on the part of *E* as to the extent or intensity of the phenomena reported resulted in a delay of the induction. Five of the *S*s remained in sensory deprivation for the allotted 6 hr. because sensory deprivation phenomena were either absent or minimal.

When *S* depressed a button to indicate eye closure, *E* used the deepening procedure suggested in the SHSS, Form B, until satisfied that he could enter

TABLE 1
ANALYSIS OF VARIANCE SUMMARY FOR THE
STANFORD HYPNOTIC SUSCEPTIBILITY
SCALE SCORES

Source of variation	df	Mean square	F
Conditions (A)	1	66.15	23.71*
Error	18	2.79	
Trials (B)	2	44.45	6.25**
A × B	2	24.91	3.50*
Error	36	7.11	
Total	59		

* $p < .05$.

** $p < .005$.

the cubicle and continue the administration of the scale. Except for minor wording changes to account for *S*'s reclining position, the scale proved applicable for administration in the deprivation cubicle itself.

Follow-up Session

Approximately 1 wk. following the experimental session, each *S* returned and again received the SHSS, Form A, under standard conditions. All *S*s in the experimental group were then interviewed in an exploratory attempt to identify those factors responsible for alteration or stabilization of susceptibility.

Control Group

A control group was included to test for the effects of repeated hypnosis upon the susceptibility of resistant *S*s and to determine the demand characteristics of the experimental procedure, minus sensory deprivation. Ten additional *S*s were selected from the same volunteer population and were assigned to the resistant control group. Methods of screening and criteria for inclusion in the sample were identical to those of the experimental group.

The *S*s were told that they were being asked to participate in an experiment measuring the effect of familiarity of surroundings on hypnotic susceptibility and that two sessions beyond the screening interview were involved. The demand characteristics of these instructions favored an increase in susceptibility. During the session corresponding to the sensory deprivation session of the experimental group, *S*s returned to the lab and sat in the deprivation cubicle, now fully lighted and equipped with magazines, radio, and comfortable chairs. The *S*s were free to study, read, sleep, or talk to *E*. More than one *S* was in the cubicle at a time, except for two *S*s who participated singly because of scheduling problems. For these two *S*s, *E* remained in the lab, thus preventing any social deprivation.

After 4½ hr. (the mean deprivation time of the experimental group), hypnotic induction was begun, using the same procedures as were used for the experimental group. The *S*s in the control group also returned for a final session to assess their final level of susceptibility using the SHSS, Form A.

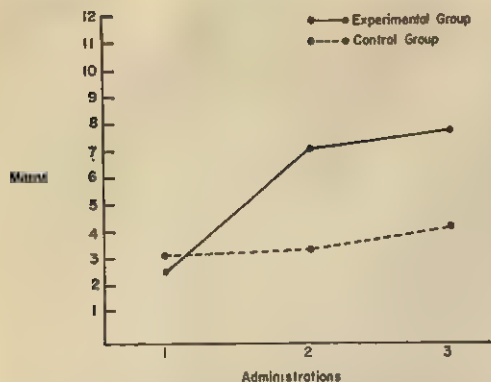


FIG. 1. Mean SHSS scores for experimental and control groups.

Dependent Variables

Four indexes of sensory deprivation were recorded for the experimental group: (a) time in deprivation—the time interval beginning with lights out and ending with the initiation of the hypnotic induction; (b) verbalization—the time *S* engaged in any type of verbalization was expressed as a percentage of the total time in deprivation; (c) stimulus hunger—stimulus-seeking behavior was obtained by counting the number of requests for stock market quotations; (d) arousal—the GSR served as a measure of *S*'s arousal both during sensory deprivation and during hypnotic induction. GSR activation was computed by dividing the number of discriminable GSRs for the period of confinement and period of induction by the time *S* remained in each condition. A discriminable GSR was operationally defined as an inflection of 1,000 ohms followed by a decline of 1,000 ohms. In addition to these objective measures, *E* recorded those cues which he used as signals to initiate hypnotic induction. This was done prior to *E*'s entry into the deprivation cubicle for the administration of the SHSS, Form B, and therefore was independent of *S*'s subsequent enhancement of susceptibility.

Scoring

To determine the reliability of *E*'s administration and scoring of the SHSS, 10 administrations of the scale were observed and scored by an independent rater. A rank-order correlation of .98 was obtained between *E*'s scoring and that of the rater. Differences in *S*'s SHSS scores between initial and final sessions were defined as enhancement scores (E scores). Previous research (Hilgard, 1965) has shown that no correction for practice effects is necessary when taking repeated measures of hypnotic susceptibility on the SHSS.

RESULTS

The SHSS scores of the experimental and control groups were subjected to an analysis

of variance (Edwards, 1960). Table 1 summarizes the results of this analysis.

The *F* ratio between conditions of administration (23.71) was significant at the .005 level, indicating that the average performance over the three administrations of the SHSS scale was significantly different for the two groups. The *F* ratio for trials (6.25) was significant at the .005 level, indicating that the three administrations of the SHSS averaged over the two groups also differed significantly. The Trials \times Conditions interaction was significant at the .05 level, indicating that the increases in susceptibility did not take the same form for the experimental and control groups. A plot of the mean increases in susceptibility for experimental and control groups is presented in Figure 1.

The means of the three administrations of the SHSS to the experimental group were compared using Duncan's new multiple-range test. Significant differences in the mean SHSS scores of the experimental group were obtained between Sessions 1 and 2 and between Sessions 1 and 3. The difference between the second and third sessions for both the experimental and control groups was not significant. The mean enhancement score of 4.80 for the experimental group differed significantly (*t* test) from the mean enhancement score of .90 for the control group at the .005 level.

The percentage of *S*s passing each of the items of the SHSS for the experimental and control groups is presented in Figures 2 and 3, respectively.

Although relatively uniform increases in

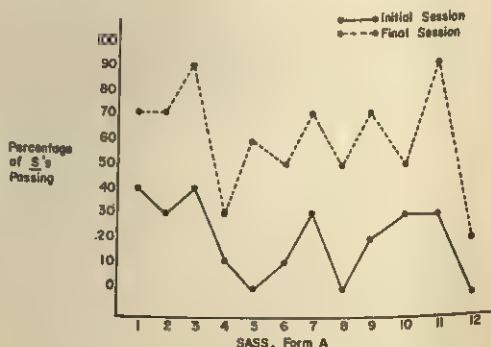


FIG. 2. Percentages of resistant experimental group passing individual items of the SHSS, Form A, for the initial and final sessions.

the percentage of *Ss* passing individual items were obtained, the shift in percentages after exposure to sensory deprivation was most pronounced for the items of finger lock, verbal inhibition, and posthypnotic suggestion.

An exploratory attempt was made to determine what sensory deprivation behaviors were predictive of enhancement of susceptibility. Spearman rank-order correlations of $-.18$, $.40$, $.56$, $-.08$, and $.19$ were obtained between enhancement scores and time in deprivation, verbalization, stimulus hunger, arousal during deprivation, and arousal during induction, respectively. None of these correlations reached statistical significance. Following the collection of data from the experimental group, *E* reviewed the tape recordings of the deprivation sessions and ranked the experimental *Ss* with respect to the intensity of their reaction to sensory deprivation. This ranking was not independent of *E*'s knowledge of *S*'s hypnotic enhancement. A nonsignificant correlation of $.54$ was obtained between *E*'s ranking and *S*'s enhancement scores.

DISCUSSION

Three alternate interpretations of the results should be considered. At the time sensory deprivation symptoms were elicited, *S* might already have been in an altered state similar to the putative one which is regnant during hypnosis. If this were the case, *S* might have been highly suggestible and the hypnotic induction procedure was actually extraneous. Or, the distress of sensory deprivation enhanced *S*'s response to any kind of instructions. Finally, the demand characteristics of sensory deprivation were greater than those of the control condition. The first two alternative interpretations are inconsistent with research reported by Zuckerman, Albright, Marks, and Miller (1962). After 6 hr. of deprivation they found that their *Ss* did not increase in suggestibility when tested for their detection threshold of a light. In the same study they found that their *Ss* manifested an inhibition of productivity when tested for free association and word naming after 3 and 6 hr. of sensory deprivation. The *Ss* were under sensory deprivation at the time of testing, and many symptoms had been produced. The problem posed by a dif-

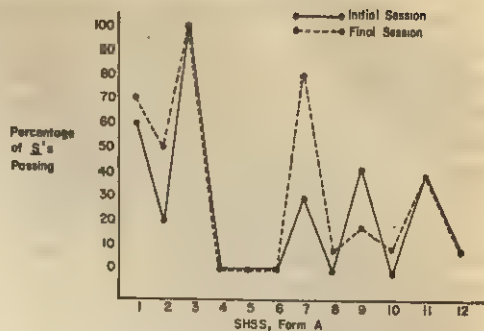


FIG. 3. Percentage of resistant control group passing individual items of the SHSS, Form A, for the initial and final sessions.

ferential in demand characteristics of the experimental and control conditions was directly tested by Zuckerman and Cohen (1964) when they increased progressively the demand characteristics for four groups using different degrees of suggestion and a placebo effect. The results were negative despite the use of a variety of dependent variables.

It should be noted also that none of these alternate interpretations readily explain the retention of the enhanced level of susceptibility 1 wk. later. According to the theory under investigation, the brain mechanisms underlying the enhanced susceptibility of the *S* have become associated with the induction procedure. In rapid inductions, a discrete cue can be used to reactivate instantaneously the brain mechanisms in question (Reyher, 1964).

In addition to its immediate relevance to the present investigation, the study by Zuckerman and Cohen (1964) is of general interest. Their results show that scores on the dependent variable are not a monotonic function of the number or saliency of cues relating to the hypotheses and expectations of the investigator. Although they did not draw this conclusion, it would seem that once *S* sizes up the relationship between the experimental design and the desired outcome of the research, his effort to please *E* is a constant value related to stable personality characteristics. Further improvement on the dependent variable can be effected only by increasing motivation through ego-involving instructions and procedures.

The results obtained here were in marked contrast to those reported by Levitt et al.

(1962), who found no increase in susceptibility as a result of exposing refractory Ss to sensory deprivation. However, a comparison of the procedures of the two studies showed marked differences. In contrast to Levitt's procedure, Ss in the present study were encouraged to report their subjective reactions to deprivation. Thus, despite the clinical nature of the decision to initiate induction, it was possible to time the onset of the hypnotic induction with the emergence of sensory deprivation phenomena.

In the present study visual stimulation was reduced to a minimal level while auditory stimulation was homogenized at near normal levels. In Peña's study both auditory and visual stimulation were homogenized at near normal levels. The most significant difference between the studies, however, was that Peña's Ss were removed from isolation prior to hypnotic induction. Because of the typically subtle and often transitory effects of sensory perceptual isolation, sudden removal from isolation is likely to have reduced the effect of his procedures. According to the theory under investigation here, the removal of an S from isolation should be accompanied by a reinstatement of higher levels of neuronal integration antagonistic to the development of successful hypnotic induction.

In contrast to Peña's control group, the control Ss in the present study produced no significant increases in susceptibility as a result of either repeated hypnotic inductions or the implied suggestions of the experimental design. In Peña's study, the group receiving $1\frac{1}{2}$ hr. of isolation did not differ significantly in enhancement from the gains of the control group. Therefore, the only clear effect that can be attributed to perceptual isolation occurred with Peña's 3-hr. group, and no attempt was made to determine the permanence of the gains in susceptibility.

The positive results of the present study are consistent with the theories of both Gill and Brenman (1959) and Reyher (1964). Gill and Brenman stated that hypnosis results from a regression of autonomous ego functioning initiated by sensory-motor deprivation occurring in a particular interpersonal relationship. The conditions imposed on Ss in sensory deprivation reproduce one of the sig-

nificant characteristics of standard hypnotic induction—the restriction of sensory input. Since these conditions initiate regressive movement in behavior, exposure to sensory deprivation should enhance hypnotic susceptibility by facilitating the regression in ego functioning necessary for successful hypnosis.

Reyher (1964) contends hypnosis is possible only after lower levels of neuronal integration become dominant in the organization of brain functioning and the regulation of behavior. The ascendancy of a lower level of neuronal integration is brought about in a susceptible S because he has abrogated an analytic-synthetic and adaptive attitude toward sensory input. He has accepted the hypnotist's assumption of this role. He listens passively to him. However, anxiety generated by the hypnotic induction because of intrapsychic conflict is sufficient to prevent this reorganization, as anxiety generates defense and other forms of adaptive behavior. Whatever the cause of resistance, people who are initially nonhypnotizable should become susceptible if a procedure such as sensory deprivation is used to bring about the necessary reorganization of brain mechanisms before standard hypnotic techniques are employed. Once an S experiences the state of hypnosis, the accompanying organization of brain mechanisms becomes associated with or conditioned to the cues inherent in the induction procedure, thus facilitating future inductions, maintaining a stable level of susceptibility, and permitting rapid hypnotic induction by posthypnotic signal.

The results of Peña's research and the procedures of the present investigation represent a promising beginning. More research is needed to clarify the relationship between specific sensory deprivation symptoms and a readiness for hypnosis. Though none of the four behavioral indicators of sensory deprivation was significantly related to enhancement scores, three of them covaried in the expected direction.

Although the theory cannot be tested directly for both technical and ethical reasons, it does have heuristic value. It consists of an integrated set of postulates about neurophysiological and psychodynamic processes which specify necessary and sufficient condi-

tions for hypnosis. The available data, including those reported here, on hypnosis, regression, and the like, are consistent with the theory; moreover, major features of neurotic and schizophrenic reactions and a variety of other manifestations of psychopathology are consistent with it.

REFERENCES

- AS, A., HILGARD, E. R., & WEITZENHOFER, A. M. An attempt at experimental modification of hypnotizability through repeated individualized hypnotic experience. *Scandinavian Journal of Psychology*, 1963, 4, 81-89.
- BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology*, 1954, 8, 70-76.
- EDWARDS, A. *Experimental design in psychological research*. (Rev. ed.) New York: Rinehart, 1960.
- GILL, M. M., & BRENNAN, M. *Hypnosis and related states: Psychoanalytic studies in regression*. New York: International University Press, 1959.
- HILGARD, E. R. *Hypnotic susceptibility*. New York: Harcourt, Brace & World, 1965.
- LEVITT, E. E., BRADY, J. P., OTTINGER, D., & HINESLEY, R. Effects of sensory restriction on hypnotizability. *Archives of General Psychiatry*, 1962, 7, 343-345.
- LONDON, P. The induction of hypnosis. In J. E. Gordon (Ed.), *Handbook of hypnosis*. New York: Macmillan, 1967.
- PEÑA, F. Perceptual isolation and hypnotic susceptibility. Unpublished doctoral dissertation, Washington State University, 1962.
- REYHER, J. Brain mechanisms, intrapsychic processes and behavior: A theory of hypnosis and psychopathology. *American Journal of Clinical Hypnosis*, 1964, 7, 107-119.
- SANDERS, R. S. The utilization of sensory deprivation procedures for increasing hypnotic susceptibility. Unpublished doctoral dissertation, Michigan State University, 1967.
- WISEMAN, R. J., & REYHER, J. Procedure utilizing dreams for deepening the hypnotic trance. *American Journal of Clinical Hypnosis*, 1962, 5, 105-110.
- ZUCKERMAN, M., ALBRIGHT, R. J., MARKS, C. F., & MILLER, G. L. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychological Monographs*, 1962, 76(30, Whole No. 549).
- ZUCKERMAN, M., & COHEN, N. Is suggestion the source of reported visual sensations in perceptual isolation? *Journal of Abnormal and Social Psychology*, 1964, 68, 655-660.

(Received January 9, 1968)

SYSTEMATIC DESENSITIZATION: EXPECTANCY CHANGE OR COUNTERCONDITIONING?¹

JAMES E. MARCIA, BARRY M. RUBIN

State University of New York at Buffalo

AND JAY S. EFRAN²

University of Rochester

Forty-four snake and spider phobic Ss, selected from a large pool of undergraduates were exposed to either (a) a form of systematic desensitization treatment, (b) a technique, called T-scope therapy, which embodies most of the expectancy-manipulating features of desensitization, but does not contain the technical elements of the procedure (i.e., relaxation, visualization, and the construction of an anxiety hierarchy), (c) T-scope therapy, presented as an "incomplete" and probably ineffective form of treatment, or (d) no treatment. There were no significant differences (on self-rating, runway, or interview measures) between the effects of the systematic desensitization procedure and T-scope therapy, although Ss receiving either of these treatments improved significantly more than those who received no treatment or T-scope therapy administered under the "low-expectancy" condition.

In general, explanations concerning the effectiveness of systematic desensitization therapy have focused on noncognitive factors (e.g., Wolpe, 1958; Wolpe & Lazarus, 1967), although the importance of cognitive factors has sometimes been implied in the writings of behavior therapists themselves. Lang (1966) pointed out that "in general, the subject has a better chance of changing his behavior if he is given meaningful feedback and encouraged to 'think about' alternatives . . . [p. 376]." Efran and Marcia (1967) have proposed a cognitive-expectancy model as an alternative way of interpreting the effects of systematic desensitization. The observed de-

creases in avoidant behavior are attributed mainly to changes in patients' expectations concerning the phobic object rather than to the conditioning of "antagonistic responses." From this point of view, patients come into treatment anticipating that contact with the phobic object will cause reactions of fear and disgust. They then go through a sensible, structured, "scientific" procedure that tends to convince them that they will be able to approach the phobic object with greater equanimity. As they become convinced, they may alter their behavior (i.e., approach the phobic object), consistent with their newly acquired expectancies. Hence treatment has been successful as a function of cognitive change rather than counterconditioning.

The present study had two objectives: (a) to demonstrate that a technique (T-scope therapy) embodying the expectancy-manipulating features of systematic desensitization, but without the technical elements of the procedure (i.e., relaxation, construction of anxiety hierarchy, and visualization), can produce significant improvement, and (b) to compare the results of this technique with those obtained with systematic desensitization.

METHOD

Because of difficulties in obtaining therapists and high-fear Ss, this study was run in two parts: T-

¹ This research was supported in part by National Institute of Mental Health Grant PN-8086 and by National Institutes of Health Biomedical and National Science Foundation institutional grants to the State University of New York. Portions of this research were presented at the 1967 American Psychological Association Convention. The authors wish to thank William Anthony, Robert Aronoff, Sheila Bob, Robert Carkhuff, James Dooling, Theodore Friel, Bernard Harris, Sandra Harris, Douglas LaBier, Michael McKenna, Marjorie Plumb, Marvin Podd, Thomas Polley, B. James Starr, and Ronna Tager who served as therapists for this study. Special thanks also to Diana Jordan for her help in coordinating the project.

² Requests for reprints should be sent to Jay S. Efran, Department of Psychology, University of Rochester, River Campus Station, Rochester, New York 14627.

scope therapy the first year and systematic desensitization the second. However, procedures for obtaining *Ss* and for measuring outcome were essentially the same, so that the study is reported as a single project.

Subject Selection and Treatment Groups

Subject selection. Approximately 2,800 introductory psychology students were administered a modified (15-item) form of the Fear Survey Schedule (FSS, Geer, 1965). Those who checked "terror" or "very much afraid" for either snakes or spiders and who were interested in participating were given a runway approach test (Lang, 1964) and an interview. The runway test, which was later readministered as a posttreatment measure, required *S*, accompanied by an *E*, to approach as closely as possible a preserved specimen of either a snake or a spider, which had been placed on a table at the end of a 15-ft. runway. Those *Ss* who could touch the specimen were excused from the study. (Over one-third of the potential *Ss* were excused on this basis—a fact which leaves the validity of FSS ratings of "terror" and "very much afraid" somewhat in question.) Forty-four *Ss* (12 males and 32 females) who were (a) not presently in treatment, (b) sufficiently fearful (according to the runway measure), and (c) willing to participate in the project were assigned to a treatment or control group.⁸ The five groups planned were: T-scope therapy (high expectancy and low expectancy), systematic desensitization (high expectancy and low expectancy), and control.

Fourteen graduate students and one faculty member in clinical or counseling psychology (10 males and 4 females) served as therapists. Except for one, they administered only T-scope therapy or systematic desensitization.

T-scope therapy. The *Ss* were met by their therapist and a white-coated GSR technician who manned a polygraph console during the eight 20-min. treatment sessions. During the first session, the therapist explained that treatment would be based upon two "well-established" psychological principles: the irrational and, hence, unconscious nature of phobias, and the suppressibility of a response by means of a mildly unpleasant stimulus. The *S* was told that phobic stimuli would be presented in a tachistoscope at a speed "too fast for the conscious mind to perceive," yet perceptible to the unconscious. These "subliminal stimuli" would evoke unconscious phobic responses which would be followed by a mildly painful electric shock. (Threshold values for pain

were established at the first session and subsequent values of electric shock were kept just below this threshold.) For each session, *S*—with GSR, EKG, and stimulation electrodes attached—sat on the edge of a chair, and looked into a tachistoscope. Although he believed he was watching subliminal stimuli, some of which were pictures of the phobic object, he was actually looking at blank cards. Each session consisted of 100 presentations, 16 of which were shocked, based on a fixed-random sequence with a mean of 6. At the end of each session, the GSR technician showed *S* a polygraph printout to illustrate his "rate of improvement." The records, which had been prepared in advance, indicated decreasing physiological reactions (increasing improvement) from the first to the eighth session with plateaus at the fifth and sixth sessions.

In order to create a low-expectancy group (Lo T), one group of T-scope therapy *Ss* was told that a "crucial element"—the phobic stimulus on the T-scope—would be missing; therefore, the phobic response would not be evoked, and neither conditioning nor improvement could be expected. This group received no polygraph feedback. High-expectancy T-scope *Ss* (Hi T) were simply administered the treatment as aforementioned.

Ten *Ss* (four males and six females) composed the Hi T group with 7 *Ss* (three males and four females) in the Lo T group.⁴

Systematic desensitization (DS). Each therapist was trained in standard systematic desensitization (DS) technique (Wolpe, 1958) by the second author. The *Ss* constructed hierarchies from a deck of 14 cards containing phobia-relevant scenes by choosing the 7 items which most closely represented his response to the phobic object, and which could most closely be incorporated into an equal interval scale of fearfulness. Hence, although all *Ss* began with the same 14 items, they were able to construct individualized hierarchies. Each *S* had two 45-min. practice sessions during which he constructed his hierarchy and was trained in deep muscle relaxation and visualization.

To control the time factor, all *Ss* spent a full session on each hierarchy item. This procedure deviated somewhat from traditional desensitization, where each *S* is allowed to proceed at his own pace. It was assumed on the basis of other studies (Davison, 1965; Lang, Lazovik, & Reynolds, 1965) that *Ss* might be expected to complete between two and three items per session, and that therefore devoting a session to just one item would allow more than sufficient time for most *Ss*. (This assumption turned out to be false in a number of instances, as indicated in the results.) An equivalence in sessions was thus established between T-scope therapy and systematic desensitization; actual time spent in therapy, although somewhat disproportionate for the two groups, was weighted in favor of systematic desensitization (see Table 3).

⁴ Three initial *Ss* in the Lo T group dropped out claiming pressure of final exams.

⁸ In the T-scope therapy portion, *Ss* were also given a 10-point "fear thermometer" (Lang, 1964), and their GSR responses to slides of neutral and phobic objects were obtained. Technical difficulties precluded analysis of the latter data and results from the fear thermometer are suspect because of the lack of independence from the runway and its susceptibility to situational demands (Efran & Marcia, 1967). Hence, these two measures were not employed in the desensitization portion of the study.

TABLE 1
SUMMARY OF IMPROVEMENT DATA

Group	Initial runway		Improvement (d)		Touch	Pickup	Inter-view rating	N
	M	SD	M	SD				
Desensitization ($N = 16$)	2.75	2.54	1.94	2.14	5	3	0 +1 +2	3 9 1
High expectancy T scope ($N = 9$)	2.78	1.71	1.78	1.71	2	1	0 +1 +2	3 4 2
Low expectancy T scope ($N = 7$)	2.86	2.14	-.14	1.77	1	1	0 +1 +2	5 1 1
Control ($N = 12$)	2.50	2.28	0	3.07	0	0	0 +1 +2	12 0 0

An attempt was made to create a low-expectancy desensitization group by telling half of the desensitization Ss in their initial interview that they would be getting only the first *half* of treatment and that the treatment could not work without both halves. However, posttreatment interviews showed that Ss promptly forgot or ignored this information, due partly to the competition of all the other new information to which they were exposed, and partly to their therapists being kept "blind," and, hence, unable to reinforce the expectancy. (This is in contrast to the successful expectancy manipulation for the T-scope groups where the therapist induced a low expectancy which was consistently reinforced by the therapeutic procedure itself; that is, S saw only a flash of light in the tachistoscope.) This group was combined with the high-expectancy group and the two collectively were referred to as systematic desensitization (DS). Since the low-expectancy DS Ss showed slightly greater improvement than the so-called high-expectancy desensitization Ss, combining the groups slightly raised the improvement scores of DS Ss. There were 19 Ss in this condition, 2 males and 17 females.

Control (C). Fourteen control Ss (3 males and 11 females) were simply tested at the beginning and at the end of the sessions of the particular treatment group with whom they were paired.

In spite of attempts to randomly assign Ss to the different treatment and control groups, when initial runway scores were compared, the four groups (Hi T, Lo T, SD, and C) were dissimilar. To equate them, one Hi T female, three DS females, and two C females were dropped from the data analysis on the basis of pretreatment scores.

Three aspects of the runway situation indicate improvement: post- minus pretreatment position on the runway; difference in number of Ss who touch the phobic object before and after treatment; and difference in number of Ss who pick up the object before and after treatment.

There are two ways of calculating change in runway position: ratio scores (Lang, 1964) based on

percentage of distance covered and conventional difference scores (pretreatment minus posttreatment). The data were analyzed both ways with nearly identical results and thus only conventional scores are reported.⁵ All runway measures were administered by Es having no knowledge of Ss' treatment groups.

During the posttreatment interviews, conducted by the authors and an advanced graduate student, each S rated his improvement along a 5-point scale: -2 = much worse, -1 = a little worse, 0 = no change, +1 = a little better, +2 = much better. These ratings were used as an additional measure of improvement. (The runway test and interview were administered by different individuals.)

RESULTS

Runway Approach Measure

Table 1 summarizes the improvement data for the four groups. One-way analyses of variance for unequal Ns were used to determine equivalence of runway starting place and significance of difference on posttreatment runway position. While starting place was nearly the same for all four groups ($F = .048$, $df = 3/40$), the groups differed significantly at the conclusion of treatment ($F = 3.75$, $df = 3/40$, $p < .05$).

The Mann-Whitney U was applied to ranked change scores between initial and final runway positions for the four groups. Results indicated that there was no significant difference between the DS and the Hi T groups, that each group improved significantly more than controls (both p 's $< .01$), that each

⁵ For the runway measures, Pearson $r = .84$, $N = 44$.

improved significantly more than the Lo T group (DS, $p < .05$; Hi T, $p < .06$), and that there was no significant difference between the Lo T group and controls.

Another way of looking at the runway data is in terms of improvement versus nonimprovement. Any positive change after treatment was considered as improvement; negative or no change was considered as "no improvement." These data were analyzed by means of Fisher exact probability tests and are presented in Table 2. Findings here were similar to those in the ranked change score analysis except that neither the DS nor Hi T group was significantly different from the Lo T group.

A third way of looking at the runway data is in terms of how many Ss could touch or pick up the phobic object at the conclusion of treatment who were not able to do so at the beginning (see Table 1). The only significant finding here was that the DS group had more Ss touching the phobic object at the conclusion of treatment than did the control group (Fisher exact probability = .04). Again, there were no significant differences between the DS and Hi T groups on these measures.

Interview Ratings

The Ss' ratings of change (from +2 to -2) appear in Table 1. No S indicated that he got worse. Fisher exact probabilities on these data for improvement (+1 plus +2) versus no improvement (0) indicated that there was no difference between the DS and Hi T conditions, that both of these groups showed significantly more improvement than controls, and, in addition, that DS Ss showed significantly more improvement than Lo T Ss.

Adequacy of Systematic Desensitization Treatment

Strict application of desensitization techniques would require that a scene be presented without S's signalling anxiety before moving on to the next hierarchy scene. However, in order to make all treatment groups equivalent in number of sessions, DS Ss were exposed to all seven hierarchy scenes, one per session, whether or not the previous scene had been successfully desensitized. More often

TABLE 2
COMPARISON OF GROUPS FOR IMPROVEMENT VERSUS
NO IMPROVEMENT ON RUNWAY MEASURE

Group	Im- prove- ment	No im- prove- ment	Comparison		
			Hi T	Lo T	C*
DS	11	5	.33	.19	< .01
Hi T	7	2		.16	< .01
Lo T	3	4			ns
C	2	10			

Note.—DS = systematic desensitization; Hi T = high expectancy T scope; Lo T = low expectancy T scope; C = control.

* Fisher exact probability (levels in cells).

than initially anticipated, Ss were presented with a scene without having been fully desensitized to a previous scene. If this decreased the effectiveness of desensitization, there should be a relationship between the methodological adequacy of treatment and Ss' improvement. Those DS Ss should show greater improvement who (a) successfully completed more sessions; or (b) completed their last successful session higher in the hierarchy; or (c) had a greater "amount" of completion across all sessions. To assess this, the runway measure of improvement was correlated with measures of these three variables. Extent of methodological adequacy (Variable c) was determined by assigning a 0 to a session in which S experienced anxiety with every visualization of the scene, a 1 if the session ended with an anxiety-free visualization of 2 sec., a 2 for a 3-sec. visualization, a 3 for a 5-sec. visualization, a 4 for a 10-sec. visualization, and a 5 for a 15-sec. visualization. Scores, averaged across sessions, ranged from 1.57 to the maximum possible of 5.00, with a mean of 3.85. Correlations of these three measures with runway improvement revealed no relationship between methodological adequacy of treatment and improvement.

An additional analysis of the relationship between methodological adequacy and improvement was made by splitting the DS group into subgroups on the basis of the proportion of sessions an S had successfully completed. If more than 50% of his sessions ended with his experiencing no anxiety during a 15-sec. visualization, he was put into the high group; with fewer than 50% of such

TABLE 3

COMPARISONS OF IMPROVEMENT IN DS, HI T, AND^a
PREVIOUS DESENSITIZATION STUDIES

Group	Approximate time in treatment ^a	Change (Lang) score ¹
DS	300 min.	.39
Hi T	190 min.	.33
Lang (1964)	720 min.	.34
Lang, Lazovik, & Reynolds (1965)	720 min.	.27

Note.—DS = systematic desensitization; Hi T = high expectancy T scope.

^a These figures include training sessions.

sessions, he was assigned to the low group. The Mann-Whitney *U* was performed comparing these groups with each other and with the controls. The high group improved significantly more ($p < .025$) when compared with controls than did the low group; however, the difference between high and low groups was not significant.

Table 3 presents improvement in terms of ratio scores (Lang, 1964) for the present study, Lang's 1964 study, and the Lang, Lazovik, and Reynolds 1965 study, permitting comparison of the results of the present study with those of previous studies. It is apparent from these figures that the effects of desensitization in the current study were equal, or superior, to those in the previous studies cited, and that they were accomplished in a shorter time. The same was true, of course, for Hi T therapy.

DISCUSSION

The first issue posed by this study, the therapeutic effectiveness of an expectancy-based therapy containing none of the theoretically crucial elements of desensitization, was resolved. T-scope patients with a high expectancy for success improved significantly more than controls on all measures, and more than the Lo T group on the runway measure. None of these Hi T patients got worse, although from a classical conditioning point of view, one might have expected such a result since if Ss happened to imagine seeing the phobic object in the tachistoscope, they would be associating this image with anxiety aroused by electric shock. Classical conditioning theory would predict that this association

would enhance the phobia; certainly it would not predict its reduction. Either Ss did not imagine the phobic object or the theory is inapplicable in this circumstance—perhaps both. Clearly, this theory, upon which systematic desensitization is based, does not account for the improvement found in this group.

A second issue concerned the comparison between Hi T therapy and systematic desensitization. The difference between the effects of these two treatments is nonsignificant in every analysis. Overall, desensitization has a slight numerical edge; but one is certainly on safe scientific and statistical grounds claiming equivalence for the two treatments.

Although the desensitization procedure used in this study deviated in form from the procedure used clinically, the outcomes were comparable to those obtained in previous studies. Moreover, comparisons of methodological adequacy with improvement suggest either that "being treated" itself is sufficient to produce improvement, regardless of number of completed hierarchy items, or, perhaps, that exposing Ss to items in the hierarchy for which they are unprepared is as efficacious as exposing them to only those for which they are "ready."

While this study does not demonstrate conclusively that desensitization is successful *because* it manipulates patients' expectancies, it certainly suggests that cognitive, as opposed to autonomic, factors are important components of the treatment. Moreover, expectancy manipulation, as opposed to classical conditioning, provides a theoretical rationale for improvement underlying T-scope therapy, systematic desensitization, and perhaps other forms of behavior therapy as well.

REFERENCES

- DAVISON, G. The influence of systematic desensitization, relaxation, and graded exposure to imaginal stimuli in the modification of phobic behavior. Unpublished doctoral dissertation, Stanford University, 1965.
- EFRAN, J. S., & MARCIA, J. E. Treatment of fear by expectancy manipulation: An exploratory investigation. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, 2, 239-240.

- GEER, J. H. The development of a scale to measure fear. *Behaviour Research and Therapy*, 1965, 4, 45-53.
- LANG, P. J. Experimental studies of desensitization psychotherapy. In J. Wolpe, A. Salter, & L. J. Reyna (Eds.), *The conditioning therapies*. New York: Holt, Rinehart & Winston, 1964.
- LANG, P. J. The transfer of treatment. *Journal of Consulting Psychology*, 1966, 30, 375-378.
- LANG, P. J., LAZOVIC, A. D., & REYNOLDS, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal and Social Psychology*, 1965, 70, 395-402.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.
- WOLPE, J., & LAZARUS, A. *Behavior therapy techniques*. Oxford, England: Pergamon Press, 1967.

(Received January 12, 1968)

Manuscripts Accepted for Publication in the
Journal of Abnormal Psychology

- Perceptual Styles and Their Correlates among Schizophrenic Patients: Carmi Schooler* and Julian Silverman: Laboratory of Socio-Environmental Studies, National Institute of Mental Health, Bethesda, Maryland 20014.
- Frequency of Psychotherapeutic Session as a Factor Affecting Outcome: Analysis of Clinical Ratings and Test Results: Christoph M. Heinicke*: Reiss-Davis Child Study Center, 9760 West Pico Boulevard, Los Angeles, California 90035.
- Cross-Cultural Comparison of Psychotic Syndromes: Maurice Lorr* and C. James Klett: Catholic University of America, Department of Psychology, Washington, D. C. 20017.
- Alcoholism and Clinical Symptoms: Allen Goss* and Thomas E. Morosko: Texas Research Institute of Mental Sciences, 1300 Moursund Avenue, Texas Medical Center, Houston, Texas 77025.
- Humans as Reinforcing Stimuli in Schizophrenic Performance: A. Sheldon Gelburd and James M. Anker*: Department of Psychology, University of Tennessee, Knoxville, Tennessee 37916.
- Preliminary Investigation of Covert Communication of Expectancies to Schizophrenics: James H. Trattner and Kenneth I. Howard*: Department of Psychology, Northwestern University, Evanston, Illinois 60201.
- Personality Correlates of the Goodman Sociosexual Adjustment Scale: Jerry Higgins*: Department of Psychology, University of California, Santa Barbara, California 93106.
- Span of Apprehension and Arousal in Schizophrenia: Herbert E. Spohn,* Paul E. Thetford, and Frederick L. Woodham: The Menninger Foundation, Topeka, Kansas 66011.
- Certain Initial Variables as Predictors of Change with Classical Psychoanalysis: H. Aronson* and Walter Weinbraub: Department of Psychology, Temple University Health Sciences Center, Philadelphia, Pennsylvania 19140.
- EEG and Rapid Eye Movements during Free Imagery and Dream Recall: Joseph Reyher* and Howard Morishige: Department of Psychology, Michigan State University, East Lansing, Michigan 48823.
- Predictability and Anxiety in Speech by Parents of Female Schizophrenics: Joseph Becker* and Paul Finkel: Department of Psychiatry, University of Washington School of Medicine, Seattle, Washington 98105.
- Effects of Task Complexity and Verbal Evaluation on the Learning of Normals and Schizophrenics: David L. Streiner: Department of Psychiatry, Hamilton Psychiatric Hospital, P. O. Box 585, Hamilton, Ontario, Canada.

* Asterisk indicates author for whom address is supplied.

CONCEPT LEARNING IN HYPERACTIVE AND NORMAL CHILDREN¹

VAIRA FREIBERGS² AND VIRGINIA I. DOUGLAS

McGill University

Concept learning in 65 hyperactive and 99 normal children was compared under three experimental conditions for two types of concepts. Both acquisition and reversal data were obtained. Contrary to current assumptions, the hyperactives showed no decrements on number, as compared to naturalistic concepts. Comparison of the clinical and normal samples showed no significant group differences under a continuous reinforcement schedule, with either a 4-sec. or an 8-sec. intertrial interval. Significant decrements in the hyperactives' performance were found under partial reinforcement, and persisted on retest sessions occurring 2 mo. later. These results suggest an explanation of the hyperactives' performance decrements in terms of attentional and motivational variables, rather than attributing them to global cognitive-conceptual impairments.

Hyperactivity is a common presenting symptom among children seen in psychiatric clinics, and refers to an excessive level of activity which is sufficiently sustained to become a serious source of complaint. Hyperactivity is usually found as the main symptom in a syndrome which also includes some or all of the following: poor motor coordination, distractibility, short attention span, and emotional instability (e.g., hypersensitivity and low frustration tolerance).

¹ This study is part of a dissertation submitted by the first author in partial fulfillment of the requirements for the degree of Doctor of Philosophy at McGill University. Portions of this paper were presented at the meeting of the Canadian Psychological Association, Vancouver, June 1965. This research was supported by United States Public Health Grant No. 293-60 to D. O. Hebb and Grant No. 298-3980 of the Medical Research Council of Canada to Virginia I. Douglas. The authors are indebted to John S. Werry, Fellow of the Royal College of Physicians (Canada), and Gabrielle Weiss, of the Department of Psychiatry, Montreal Children's Hospital, whose collaboration made this research possible. Grateful thanks are also expressed to the Department of Special Education, Greater Montreal Protestant School Board, and to the following school principals and their staff for their kind cooperation: G. Taylor, Oak Ridge School, Baie d'Urfée; A. Pitcairn, Willingdon School; W. Kydd, Guy-Drummond School; M. T. Craig, Westmount Park School; and J. V. Pepler, Logan School.

² Requests for reprints should be sent to Vaira Freibergs, who is now at the Université de Montréal, Institut de Psychologie, Case Postale 6128, Montreal 101, Canada.

Most of the information on hyperactivity currently available deals with questions of diagnosis and treatment. Little seems to be known of the intellectual functioning of hyperactive children, apart from the frequent observation that their school performance is not satisfactory (Clements & Peters, 1962; Laufer, Denhoff, & Riverside, 1957; Laufer, Denhoff, & Solomons, 1957). In one study (Burks, 1960), the IQs of a sample of hyperactive children were compared with those of a normal control group, but no significant differences were found.

Conceptual difficulties as characteristic of hyperactive children have been reported by several authors (Burks, 1960; Clements & Peters, 1962; Rosenfeld & Bradley, 1948) without offering any supportive objective evidence. Rather, the children's poor academic performance is somewhat freely interpreted as resulting from specific learning defects in reading, spelling, or, in some cases, only arithmetic and number concepts. Burks (1960) has suggested that the conceptual difficulties of the hyperactive reflect "inefficient patterning and processing capabilities of the brain." Similarly, Clements and Peters (1962) refer to an impaired "capacity to receive, hold, scan, and selectively screen out stimuli in sequential order."

Generalizations of this nature, while unwarranted by the mere evidence of unsatisfactory school progress, become more under-

standable in light of the widely held belief that the hyperactive syndrome is due to some form of brain damage or dysfunction. Loss of conceptual ability or "abstract attitude" is of course the classical symptom reported in individuals known to have suffered cerebral lesions (Battersby, Krieger, Pollack, & Bender, 1953; Goldstein & Gelb, 1918). A critical evaluation of the literature, however, reveals that the evidence linking hyperactivity to various signs of brain damage is extremely inconclusive (Freibergs, 1965). Similarly, a preliminary investigation of the same population from which Ss in the present study were drawn suggested brain damage as only one of several possible causes of hyperactivity (Werry, Weiss, & Douglas, 1964). Most important, the nature of the cognitive impairments, if any, should be established independently, by means of objective studies, whatever the etiological nature of hyperactivity.

The aim of the present study was to provide some objective evidence on this question by systematically comparing the performance of hyperactive and normal children in a controlled learning situation involving concept formation tasks. Long series of different stimuli were used, eliminating the possibility of correct responding by rote memorization of the correct stimuli. In order to discover the correct concept, Ss would first need the ability for abstraction in the literal sense, that is, the ability to single out as relevant particular dimensions or attributes of the stimulus complex. In addition, a symbolic categorization response would be required (e.g., the verbal label "flower") which would help in choosing the correct pictures despite considerable variation in their physical appearance. Finally, the task required the ability to modify responses on the basis of the informative feedback provided by concrete reinforcements for correct responses.

Two types of concept problems were used in this study, in order to verify the hypothesis that hyperactive children have more difficulty with number concepts. Variations were also introduced in schedule of reinforcement and length of intertrial interval in an attempt to manipulate the difficulty level of the tasks and increase the likelihood of detecting any deficits that may be present.

METHOD

Subjects

Hyperactive sample. The 65 Ss included in the hyperactive group were selected by two staff psychiatrists on the basis of criteria designed to make the sample as homogeneous as possible. Hyperactivity had to be the major presenting symptom, being chronic, sustained, and reported by both the parents and the school. Any Ss diagnosed as defective, psychotic, epileptic, or primarily neurotic were excluded, as were those with definite indications of brain damage. The EEG data and early medical history for part of this sample have been presented by Werry, Weiss, and Douglas (1964). The Ss were receiving neither drugs nor psychotherapy at or immediately preceding the time of assessment. They ranged in age 6-12 yr., with a mean age of 8.6. Their mean IQ on the WISC was 103.8, with a range of 83-127. All but three of the clinical Ss were boys.

Normal sample. The 99 control Ss were students from schools in the Montreal area who had no indications or known history of behavioral problems or emotional disturbance. The normal children ranged in age 6-12 yr., with a mean of 8.9. Since individual IQ measures were not available for these Ss, the Vocabulary subtest of the WISC was administered and used as an IQ estimate. Their mean scaled score on the Vocabulary subtest was 12, $SD = 1.7$. The mean Vocabulary score for the hyperactive Ss was 11.3, $SD = 2.8$. The difference between the two overall samples was nonsignificant on a two-tailed test ($t = 1.80$, $df = 157$). All but 3 of the normal Ss were boys.

Apparatus

The apparatus consisted of a $12 \times 12 \times 18$ in. gray metal cabinet which allowed for presentation of two contiguous $2\frac{1}{2}$ -in. square stimulus pictures on a rear-projection Plexiglas screen. A response key was located just below each stimulus. Marble rewards for correct responses were released through an opening immediately upon depression of the response key.

Stimuli

Four different concept problems, of a type used by Osler in previous studies (e.g., Osler & Weiss, 1962), were used. Each concept was represented by a set of 150 unduplicated pairs of stimuli, one of which was always an exemplar of the concept. Two object concepts, "flower" and "bird," were represented by naturalistic colored pictures. The other two problem sets represented the number concept "two" by means of black figures of different geometric shapes and dots of different colors, respectively.

Procedure

Each S was tested individually in four separate sessions. Instructions were given about the nature of the task and rate of reinforcement to be expected (continuous versus intermittent). The Ss were tested to a criterion of 10 consecutive correct responses for

TABLE 1
MEAN TIME INTERVALS IN DAYS SEPARATING
DIFFERENT TEST SESSIONS FOR EACH
EXPERIMENTAL SUBGROUP

Treatment condition	First to second session		Test-retest interval	
	Pretest	Retest	Object concept	Number concept
Normal Ss				
Continuous	6.9	4.1	53.4	50.6
Partial	6.8	4.1	53.0	50.3
Delay	4.0	—	—	—
Hyperactive Ss				
Continuous	6.6	3.2	66.4	62.5
Partial	5.0	4.8	68.8	68.1
Delay	3.0	—	—	—

a maximum of 300 trials. Whenever an S reached criterion in less than 300 trials, he was switched to a reversal stage of learning without any warning. Exemplars of the concept just learned (e.g., flowers) were no longer reinforced, but reinforcements could be obtained by choosing the nonexemplars of the concept. Testing was again continued until a reversal criterion of 10 consecutive correct responses or for a maximum of 300 trials since the beginning of testing.

Design

The basic design of the study was a $2 \times 3 \times 2$ factorial with replication on the third variable. The two types of Ss, hyperactives and normal controls, were randomly assigned to either of three treatment conditions, continuous (CR), partial (PR), or delay (DL), with the restriction that age and average IQ estimate be kept comparable among the groups.

The Ss in the CR groups received 100% reinforcement for correct choices in a two-choice discrimination situation, with a 4-sec. intertrial (postreinforcement) interval. The PR groups received 50% reinforcement on a fixed ratio schedule (every second correct response was reinforced), also with a 4-sec. intertrial interval. The Ss in the DL groups were on a continuous reinforcement schedule, but had an intertrial interval of 8 sec.

The first two test sessions constituted the pretest phase of the study, with all Ss receiving the flower concept, followed by the two black shapes concept. Retest took place about 2 mo. later, with all Ss receiving the bird concept, followed by the two colored dots concept. The average time intervals between test sessions are given in Table 1. The hyperactive Ss were placed on out-patient drug therapy immediately after initial assessment, and were taking either chlorpromazine or a placebo at the time of retesting (see Freibergs, Douglas, & Weiss, 1968; Werry, Weiss, Douglas, & Martin, 1966).

Additional control samples. In the basic design of this study, the effects of order of presentation of the two concepts are confounded with possible differences

in their level of difficulty. Counterbalancing at the problem level, however, was not attempted at the outset, because of the limitations in the number of hyperactive Ss available. On the basis of a preliminary analysis of the data, two additional control groups were added at the CR level. This "order control" sample consisted of 22 normal and 10 hyperactive Ss who received the two pretest problems in a reversed order: two black shapes, followed by flowers, but were not tested in the retest phase.

RESULTS

Pretest Data

Response measures. Performance on each concept was scored in terms of the number of trials required to reach criterion. The Ss who did not reach criterion within the maximum of 300 trials were assigned an arbitrary score of 300.

The distribution of solvers (Ss reaching criterion) and of nonsolvers for each subgroup is given in Table 2. There were significantly more solvers among the normal children ($\chi^2 = 14.55$, $df = 1$, $p < .001$) as well as increasing numbers of solvers in the PR, DL, and CR subgroups ($\chi^2 = 17.62$, $df = 2$, $p < .001$). The two types of concept problems did not yield significant differences in the distribution of solvers and nonsolvers.

Basic experiment. Means and standard deviations for the trials to criterion measure are presented in Table 3. Because of heterogeneity in the group variances, all statistical calculations were performed after a square-root transformation of the data. A fixed factor model analysis of variance was employed

TABLE 2
NUMBER OF SUBJECTS PER GROUP WHO REACHED
CRITERION PERFORMANCE ON
PRETEST CONCEPTS

Category	Normals			Hyperactives		
	Continuous	Partial	Delay	Continuous	Partial	Delay
Flower concept						
Solvers	26	22	25	19	7	11
Nonsolvers	0	3	1	1	13	4
Total	26	25	26	20	20	15
Two-shapes concept						
Solvers	25	20	23	20	10	13
Nonsolvers	1	5	3	0	10	2
Total	26	25	26	20	20	15

($2 \times 3 \times 2$ factorial with repeated measurements on the last factor), with an unweighted-means correction for unequal cell frequencies (Winer, 1962, pp. 337, 342).

The main effect for *S* classification was significant at the 5% level. Specific normal-hyperactive comparisons, collapsing data across problems, showed that this effect was due mainly to the hyperactives doing more poorly than the control children under partial reinforcement ($F = 3.65$, $df = 1/126$, $p < .05$). The difference between the hyperactive and the control children was not significant under continuous reinforcement, that is, in either the CR ($F = .09$, $df = 1/126$) or in the DL condition ($F = 2.80$, $df = 1/126$).

The main effect for experimental treatment was significant at well beyond the 1% level of confidence. Specific paired comparisons showed that *Ss* receiving 100% reinforcement for correct responses performed significantly better than those receiving 50% reinforcement ($F = 45.87$ for CR versus PR; $F = 13.42$ for DL versus PR, $df = 1/126$, $p < .01$ in both cases). Concept learning under PR thus is considerably more difficult for all *Ss*, and particularly difficult for the hyperactive *Ss*. No systematic effects due to intertrial interval were found, the CR versus DL comparison being nonsignificant for both types of *Ss*. There was a trend for poorer performance on the first problem by the hyperactive DL group, especially for the younger *Ss*, but this trend did not reach significance.

The overall difference between the two problems (i.e., the two test sessions) was significant at the 5% level and there was also

TABLE 4

MEAN TRIALS TO CRITERION FOR ORIGINAL AND ORDER CONTROL GROUPS UNDER CONTINUOUS REINFORCEMENT

<i>S</i>	Original continuous group		Order control group	
	First problem flower	Second problem two shapes	First problem two shapes	Second problem flower
Hyperactive	134.4	51.7	99.4	36.4
<i>N</i>	13	13	10	10
Normal	87.6	89.1	113.9	42.4
<i>N</i>	26	26	22	22

a significant *Ss* \times Problems interaction. This reflects the fact that hyperactive *Ss* in the CR and DL groups showed a significant improvement on the second problem, with a slight trend in that direction by the hyperactive PR group. All normal *Ss* performed at a comparable level on both pretest problems.

Order control data. In order to evaluate the respective effects of problem type and of the order of presentation (practice effect), the data of the original CR groups and those of the order control samples were combined into a Latin-square design as shown in Table 4. Separate analyses were done for the hyperactive and the normal *Ss*. Since the 10 additional hyperactive *Ss* obtained ranged in age 7-9 yr., only the 7-, 8-, and 9-yr.-old *Ss* of the original hyperactive CR group (a total of 13 *Ss*) were included in this analysis.

Results of the Latin-square analyses showed that, first of all, the differences between the two separate groups were nonsignificant, indicating that the original and order control samples are quite comparable. For the normal samples, the effects of order of presentation and type of concept problem were both significant at the 5% level of confidence. The normal *Ss* thus showed positive transfer to the second problem, but this was obscured in the original CR sample by the fact that the two concept is apparently more difficult for them than the flower concept. In the hyperactive data, on the other hand, the differences due to order of presentation were highly significant ($F = 12.91$, $p < .01$), but there were no significant differences due to the type of concept ($F = .24$). Thus, under CR, hyper-

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR TRIALS TO CRITERION ON THE TWO PRETEST CONCEPTS

Measure	Normals			Hyperactives		
	Continuous	Partial	Delay	Continuous	Partial	Delay
Flower concept						
<i>M</i>	87.6	162.6	98.0	122.3	221.4	169.3
<i>SD</i>	61.8	86.1	92.0	87.7	113.7	122.2
Two-shapes concept						
<i>M</i>	89.1	154.5	95.4	47.4	201.9	104.7
<i>SD</i>	73.2	100.9	98.0	37.7	110.2	102.8

active Ss show a large amount of positive transfer which is independent of the particular concepts involved.

Retest Data

The main original aim in gathering the retest data had been to evaluate the effects of chlorpromazine on learning in the hyperactive children. Since drug-placebo comparisons had revealed no systematic effects attributable to type of medication received (see Freibergs et al., 1968), the data for all hyperactive Ss were combined, giving a sample of 19 Ss under CR and 18 under PR. The combined data were then used for comparisons with the normal CR and PR subgroups. Although the hyperactives were retested an average of 2 wk. later than the controls, this seems unlikely to bias the results since correlations between test-retest time intervals and equivalent difference scores were not significant.

The retest data were analyzed by means of an analysis of covariance, with the pretest scores on the equivalent concept serving as covariates. Separate calculations were made for all Ss under CR and under PR. The results showed no significant differences between hyperactive and normal Ss in the CR condition ($F = .99$, $df = 1/42$), while under PR the difference was significant at the 1% level ($F = 8.00$, $df = 1/40$). The pretest deficit shown by the hyperactives under PR was not only maintained over the retest sessions, but actually became more evident. While the control Ss under PR improved significantly on the retest sessions, the hyperactive PR Ss showed no apparent improvement over their pretest performance. The effects of problem type were not significant.

Reversal Data

Since only Ss who had reached the acquisition criterion could be meaningfully placed on a reversal schedule, the possibility of bias should be considered in the evaluation of these data. Analyses of covariance were used because of the possibility of a systematic relationship between the number of acquisition and reversal trials. Separate analyses were made for each of the four concepts, with adjustments for unequal cell frequencies.

Considering first the pretest data, neither

the main effects, nor the interaction were significant in either analysis. Thus, with number of acquisition trials controlled, the hyperactive Ss showed no indications of "rigidity," or difficulty in switching responses under reversed reinforcement contingencies. Similarly, the differences between the CR and the PR groups were nonsignificant, giving no indication of an equivalent to the partial reinforcement effect (PRE) at this stage of testing.

Summaries of the analyses of covariance for both retest concepts are presented in Table 5. Here both main effects were significant, more trials to reach reversal criterion being required by the hyperactive, as compared to the normal, children and by Ss under PR as compared to those under CR. The interaction term is also highly significant on the second concept, reflecting the fact that the differences between hyperactive and normal Ss were markedly larger under the PR condition.

DISCUSSION

The main aim of the present study was to gather objective evidence on the presence or absence of cognitive deficits of a conceptual nature in hyperactive children. The first main finding is that hyperactives were just as efficient as the normal controls in discovering the correct concept in a standard (CR) concept-learning situation. This argues quite strongly against the currently held opinion that they suffer from basic deficits in the fundamental cognitive operations involved in the concept formation process, namely, abstraction of stimulus dimensions, production of symbolic categorization responses, and modification of response strategies on the basis of informative feedback.

The experimental situation, it must be noted, had several features which would be expected to facilitate learning by helping to arouse interest, focus attention, and reduce boredom. None of these should have been of any help, however, had the hyperactive Ss been truly unable to function at a conceptual level.

One feature of the experimental situation which may have made a difference is the fact that the task was self-paced, since stimuli remained in view until S had made his response.

TABLE 5
ANALYSES OF COVARIANCE OF RETEST REVERSAL SCORES WITH
EQUIVALENT ACQUISITION SCORES AS A COVARIATE

Concept problem	Source of variation	df	MS	F
Birds (first retest session)	Groups (normal versus hyperactive)	1	173.75	5.37*
	Reinforcement (CR versus PR)	1	147.93	4.57*
	Groups \times Reinforcement	1	57.43	1.78
	Residual	76	32.34	
Two colored dots (second retest session)	Groups (normal versus hyperactive)	1	296.71	22.79**
	Reinforcement (CR versus PR)	1	154.32	11.85**
	Groups \times Reinforcement	1	109.07	8.38**
	Residual	71	13.02	

Note.—CR = continuous reinforcement; PR = partial reinforcement.

* $p < .05$.

** $p < .01$.

With stimuli presented for a fixed period of time and paced at a predetermined rate, there is a possibility that hyperactive Ss might do more poorly than they did here.

The general assumption that hyperactive children have particular difficulty with arithmetic and number concepts is also contradicted by the present data. There was no indication that hyperactives had any difficulty with the two number concepts included in the present study. Indeed the contrary was found, since for the normal children the two black shapes problem was significantly more difficult than the flower problem. The hyperactive children did significantly better on the second test session than on the first, regardless of which concept was presented. The reason for the discrepancy between the present findings and previous literature seems simply due to a different interpretation of what is meant by "number concepts." So far as can be ascertained, the difficulty with number concepts has been inferred on the basis of poor arithmetic performance in school. Traditional primary school arithmetic, however, deals mainly with the rote, mechanical, application of memorized information such as tables of multiplication, rules for "carrying," etc., and typical tasks (addition, long division) do not require any particular understanding of concepts of numerosity and the like. What they do require is concentrated and sustained attention, qualities in which a hyperactive child may indeed be lacking.

A further discrepancy with general descriptions of hyperactive children arises with respect to the question of perseveration. No

doubt because of the belief in an organic etiology, hyperactives have been said to be more stimulus-bound and to perseverate with the same response, much as brain-injured patients are known to do. In the present study this hypothesis was tested in two different ways. First, the presence of any primitive position habits was checked by analyzing sequences of left-right response choices for all Ss (see Attneave, 1959, p. 24 for the method). A comparison of the mean results (not reported here) showed no indication of systematic position preferences by either group.

The other measure of perseveration was given by the reversal data. Here again, on the two pretest sessions, the hyperactive Ss showed no tendency to persevere with the previously correct response when the reinforcement contingencies were reversed after reaching criterion. In view of this, the significant differences which appeared in the retest reversal data cannot be taken as evidence of perseveration, but must be explained in terms of the differences in test-retest transfer that they represent.

The second main finding is that under partial reinforcement, the hyperactives not only showed significant decrements on the two pretest sessions, but also failed to improve on the retest sessions, while all other Ss (including hyperactives under CR) showed considerable positive transfer.

The crucial, defining event of the PR situation is the occurrence of nonreinforcements (NR) for correct as well as incorrect responses. The best strategy for solution under PR thus would be to memorize the character-

istics of the reinforced pictures and to search for possible commonalities between them.

Considering the greater time interval which separates successive reinforced responses under PR, such a strategy could be more difficult for Ss with more rapidly fading short-term memory, or with a difficulty in rapid coding of information for memory storage. In this case, however, some decrements would have been expected for the hyperactive DL group as well, where the postreinforcement interval was twice as long as in the CR condition. While there was a tendency in that direction, no systematic group differences were obtained.

Another way of considering the PR situation is in terms of Amsel's theory (1958, 1962), according to which the occurrence of NR under PR produces frustration, a primary aversive motivational condition. While frustration theory is based strictly on runway experiments using rats as Ss, the frustrative effects of NR have also been demonstrated in token-reward situations with children (Kendler, Kendler, Pliskoff, & D'Amato, 1958; Lambert, Lambert, & Watson, 1953; Longstreth, 1960). According to Amsel (1962), a secondary form of frustration, r_F , or fractional anticipatory frustration, develops over a series of learning trials through a process of classical conditioning. This, in turn, produces two different effects which are in competition during the intermediate phase of learning: activating, or drive effects, and inhibitory effects producing a partial decrease in strength of the instrumental response.

On the basis of the above formulation, the performance decrements of hyperactive Ss under PR could be explained as being due to a higher than average sensitivity to the frustrative effects of NR. The inhibitory component of r_F would favor abandoning any hypothesis that had not resulted in reinforcement. Discarding hypotheses after one, or even a few, NR would make it impossible to reach the correct solution under PR. Similarly, the activation component of r_F , by raising Ss' level of arousal above an optimum limit, would decrease the likelihood of task-relevant discriminations by interfering with the cue function of stimuli (Hebb, 1955). This would retard the discovery of the correct concept by

interfering with the search for elements common to all reinforced pictures.

That many hyperactives under PR found it difficult to continue searching for hypotheses can be seen from the comments noted in their test records. Some children clearly abandoned any attempts at rational solution, saying "the rule changes all the time," or "there is no rule." A few others claimed they had "the right answer," despite the absence of a confirmation from *E*. These reactions, incidentally, are almost classical examples of denial, one of the defense mechanisms used in attempting to deal with frustration (Douglas, 1965).

This motivational explanation of the PR deficits of hyperactive Ss is consistent with the absence of any truly cognitive deficits suggested by their "normal" performance in the CR and DL conditions. It may also help to explain the retest reversal results. The frustration involved in abandoning a solution as soon as it has proved successful might have prevented the hyperactives (and all PR Ss) from taking maximum advantage of their discovery of the reversal rule during pretest. Finally, this explanation is consistent with what appears to be the fundamental difficulty underlying the hyperactives' symptoms. Distractibility, hypersensitivity, and low frustration tolerance (cf. Werry, Weiss, & Douglas, 1964) can be seen as all stemming from the same fundamental difficulty—a chronically excessive level of arousal. This causes the hyperactive child to have abnormally low response thresholds, whether the response be skeletal (hyperactivity), autonomic (irritability, hypersensitivity), or cognitive (distractibility).

The notion of low response thresholds leading to response competition helps to explain the apparently crucial importance of CR for effective concept learning in the hyperactive. At the beginning of a test session, especially the first one, the hyperactive child typically reacts to a wider range of stimuli than a control, as indicated by the observations noted during testing. The hyperactive children's overt reactions to *E*, the apparatus, objects in the room, or their own bodies would obviously compete with their attention to the task at hand. Normal children, on the other

hand, were remarkably quiet and businesslike during testing. On each reinforced trial under CR, however, the hyperactive child's attention is drawn back to the task by the immediate appearance of a visible, audible, and tangible reinforcement. Thus the advantage of the CR schedule is that it serves to reinforce task orientation (i.e., attention) in the hyperactive S, and acts as a continuing counterforce to his distractibility and physical restlessness. In addition, the positive motivational aspect of reinforcement helps to avoid the appearance of maladaptive responses such as overanxiousness or giving up.

The present findings have implications for the schooling of hyperactive children. So far, the best recommendation that could be made in severe cases was for a special environment with outside stimulation reduced to a minimum (e.g., Strauss & Lehtinen, 1947). According to the findings of the present study, favorable results may also be obtained by providing positive reinforcements at a continuous rate. Thus learning materials programmed so as to minimize the pupil's opportunities for error could be used for school subjects which create the most difficulty (spelling, arithmetic). An important proviso is that the situation should provide as much immediate sensory feedback as possible. This could be implemented, for example, by having children work at computer consoles linked to a real time system with central program storage.

REFERENCES

- AMSEL, A. The role of frustrative nonreward in non-continuous reward situations. *Psychological Bulletin*, 1958, 55, 102-119.
- AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning. Some recent history and a theoretical extension. *Psychological Review*, 1962, 69, 306-328.
- ATTNEAVE, F. *Applications of information theory to psychology: A summary of basic concepts, method and results*. New York: Holt, 1959.
- BATTERSBY, W. S., KRIEGER, H. P., POLLACK, M., & BENDER, M. B. Figure-ground discrimination and the "abstract attitude" in patients with cerebral neoplasms. *A.M.A. Archives of Neurology and Psychiatry*, 1953, 70, 703-712.
- BURKS, H. F. The hyperkinetic child. *Exceptional Children*, 1960, 27, 18-26.
- CLEMENTS, S. D., & PETERS, J. E. Minimal brain dysfunction in the school-aged child. *Archives of General Psychiatry*, 1962, 6, 185-197.
- DOUGLAS, V. I. Children's responses to frustration: A developmental study. *Canadian Journal of Psychology*, 1965, 19, 161-171.
- FEIBERGS, V. Concept learning in normal and hyperactive children. Unpublished doctoral dissertation, McGill University, 1965.
- FEIBERGS, V., DOUGLAS, V. I., & WEISS, G. The effect of chlorpromazine on concept learning in hyperactive children under two conditions of reinforcement. *Psychopharmacologia (Berlin)*, 1968, 13, 299-310.
- GOLDSTEIN, K., & GELB, A. Zur Psychologie des optischen Wahrnehmungs und Erkennungs-vorganges. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 1918, 41, 1.
- HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 1955, 62, 243-254.
- KENDLER, H. H., KENDLER, T. S., PLISKOFF, S. S., & D'AMATO, M. F. Inferential behavior in children: I. The influence of reinforcement and incentive motivation. *Journal of Experimental Psychology*, 1958, 55, 207-212.
- LAMBERT, W. M., LAMBERT, E. C., & WATSON, P. D. Acquisition and extinction of an instrumental response sequence to the token-reward situation. *Journal of Experimental Psychology*, 1953, 45, 321-326.
- LAUFER, M. W., DENHOFF, E., & RIVERSIDE, R. I. Hyperkinetic behavior syndrome in children. *Journal of Pediatrics*, 1957, 50, 463-474.
- LAUFER, M. W., DENHOFF, E., & SOLOMONS, G. Hyperkinetic impulse disorder in children's behavior problems. *Psychosomatic Medicine*, 1957, 19, 38-49.
- LONGSTRETH, L. E. The relationship between expectations and frustration in children. *Child Development*, 1960, 31, 667-671.
- OSLER, S. F., & WEISS, S. R. Studies in concept attainment: III. Effect of instructions at two levels of intelligence. *Journal of Experimental Psychology*, 1962, 63, 528-533.
- ROSENFELD, G. B., & BRADLEY, C. Childhood behavior sequelae of asphyxia in infancy with special reference to pertussis and asphyxia neonatorum. *Pediatrics*, 1948, 2, 74-84.
- STRAUSS, A. A., & LEHTINEN, L. E. *Psychopathology and education of the brain-injured child*. New York: Grune & Stratton, 1947.
- WERRY, J. S., WEISS, G., & DOUGLAS, V. I. Studies on the hyperactive child I: Some preliminary findings. *Canadian Psychiatric Association Journal*, 1964, 9, 120-130.
- WERRY, J. S., & WEISS, G., DOUGLAS, V., & MARTIN, J. Studies on the hyperactive child. III. The effect of chlorpromazine upon behavior and learning ability. *Journal of the American Academy of Child Psychiatry*, 1966, 5, 292-312.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received January 22, 1968)

INFLUENCE OF THE EXPERIMENTER ON PUPILLARY DILATION TO SEXUALLY PROVOCATIVE PICTURES

LOREN J. CHAPMAN,¹ JEAN P. CHAPMAN

University of Wisconsin

AND TERRY BRELJE²

Illinois Security Hospital

Two *E*s with markedly different personality characteristics and styles of interacting with *S*s tested male undergraduates ($N = 22$ and 25) for pupillary dilation to pictures of nude and partially clothed women, as well as to pictures of men, and to control pictures. The *S*s of *E*₁, an aloof and businesslike graduate student, dilated equally to male and female pictures. The *S*s of *E*₂, a casual outgoing undergraduate, dilated more to pictures of women than to those of men. These results indicate that pupillary dilation to sexually provocative stimuli reflect, in part, *S*'s relationship to the examiner and his view concerning permissible emotional responses to the stimuli.

As Loewenfeld (1966) has recently pointed out, research workers have known for almost two centuries that emotional and cognitive arousal produces pupillary dilation. Nevertheless, Hess and his co-workers (Hess, 1965; Hess & Polt, 1960, 1964, 1966; Hess, Seltzer, & Shlien, 1965) are responsible for the recent upsurge of interest in the effects of psychological events on pupil size. These investigators have reported, among other things, that the pupil of the human eye changes in size as a function of either the observer's interest or his positive evaluation of the stimuli that he observes. Hess and Polt (1960) reported that the pupils of 4 male *S*s were larger when *S*s observed pictures of nude women than when they viewed either control slides or pictures of nude men. Nunnally, Knott, Duchnowski, and Parker (1967), using a group of 26 *S*s, have confirmed Hess and Polt's finding that dilation occurs in males as a response to pictures of seminude females. Hess, Seltzer, and Shlien (1965) have further reported "pilot" data indicating that pupillary response may

be a useful measure of heterosexual and homosexual interests. These investigators reported that they used as their heterosexual *S*s five "Students or workers in our laboratory . . . well known to us over a period of several years [p. 165]." Their homosexual *S*s were five other students who admitted their homosexual activity. The investigators used each *S*'s pupil size in response to nonsexual art as a base line from which to compute for him a score of change in pupil size in response to female pictures and to male pictures, as well as a score of the difference between these two scores. Hess et al. found no overlap between their heterosexual and homosexual *S*s on this difference score, and they concluded, "In the study of some aspects of personality . . . this technique appears to us to open up entirely new dimensions [p. 168]."

Enthusiasm for the promise of the technique should be tempered by the evidence from other psychological research that *S*'s response in an experiment or assessment situation may be influenced by his perception of his relationship to the examiner, and his anticipation of what the examiner expects him to do. Howes and Solomon (1950) made these points dramatically in their now classic critique of McGinies' (1949) study of perceptual defense against taboo words. Masling (1960) reviewed a number of studies that demonstrate such effects in psychodiagnostic testing. Recently Rosenthal (1966) has marshaled considerable

¹ Requests for reprints should be sent to Loren J. Chapman, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

² The authors are indebted to Eckhard Hess and James Polt for manufacturing the stimulus slides used in the present study and for assistance and advice in copying their equipment. The authors are also indebted to Neil Tafeen, Thomas Stern, and Lynn Kahn for assistance in conducting the study, and to Richard McFall and Rafael Klorman for their critiques of an earlier draft of the manuscript.

evidence that the expectations of *E* concerning the behavior of his *Ss* may influence their responses. Contemporary investigators of the pupillary response have ignored the possibility that *E* may influence the outcomes of his experiments.

In the course of a pilot study on pupillary response to sexually arousing pictures, the present writers had noticed that two *Es* working on the study were obtaining different results. Therefore, a study was designed to investigate more systematically the effects on the pupillary response of *E*'s personality and his style of interacting with his *Ss*.

METHOD

Experimenters

The two *Es* each tested half the *Ss* in a study of pupillary response to sexually provocative pictures. The design was such that the data of the two *Es* could be treated as two independent but parallel studies. Each *E* was told to behave in his accustomed way in his interaction with his *Ss*, but in presenting the experiment, to follow the written instructions to *S* precisely. The following description of the two *Es* represents a consensus of several observers.

*E*₁ was a 28-yr.-old married male graduate student. He wore a coat and tie and maintained with his *Ss* a businesslike, somewhat formal, and aloof relationship. His manner appeared to communicate the message that the experiment was a serious affair in which *Ss* were expected to work carefully and behave properly, much as they are expected to conform and behave properly in a classroom situation.

*E*₂ was a 21-yr.-old male unmarried undergraduate who came to the experiment dressed in khaki trousers and either a sport shirt or pullover sweater. He was an energetic, buoyant, casual young man with a very friendly but "breezy" approach to interactions with others. His style of deportment communicated fairly immediately an image of a carefree, fun-loving undergraduate, with a lively interest in the opposite sex. He interacted with his *Ss* as equals, rather than as a superior in the academic hierarchy.

The authors believe that these two styles of interaction with *S* developed differing expectations by *Ss* concerning the study, and differing views as to permissible emotional responses to the stimuli presented in it. *E*₂ established an atmosphere which, the authors believe, permitted, or even encouraged, his male *Ss* to partake freely of the enjoyment of viewing sexually provocative pictures of women. *E*₁, on the other hand, probably developed an atmosphere for the study in which his *Ss* would not be likely to view such a response as appropriate.

The two *Es* alternated in running blocks of *Ss* so as to avoid any possible changes in pupillary response as a function of the time of day or the stage in the experiment in which *S* was tested.

Apparatus

The apparatus was a near duplicate of the apparatus used by Hess and his collaborators and described by Hess (1965) and by Hess and Polt (1966). This apparatus produced photographs of the pupil of one eye on infrared film, two frames per second, while *S* viewed for 10 sec. an image projected onto the back of a translucent screen. For each slide, the total incident light at *S*'s eye was approximately 169 ftc. Since sudden sounds produce pupillary dilation (Loewenfeld, 1958), Hess' apparatus was modified by placing the projector and all relays in an adjacent room and by projecting through a double-pane glass window so that the noise of both the projector and the relays was not audible in the testing room. In an additional modification of Hess' apparatus, the film was marked when the stimulus slide was changed by directing a light beam into the focus lens of the camera. This made it possible to determine with certainty which slide was being viewed for each photograph of the pupil.

The photographs of the pupil were projected and magnified on a ground glass screen mounted in the top of a table. A research assistant measured with a ruler the diameter of the projected image of each pupil. This assistant was not aware which *S*'s pupils he was measuring or which slide *S* had been viewing.

The first two and last three frames of each series of 20 photographs were omitted from the analysis of data to reduce effects from preceding slides, the effects of the momentary change in illumination when slides were changed, and any effects of *Ss* anticipating the change of slides near the end of a series. Also omitted were any photographs on which *S* blinked or moved his eyes. A mean diameter was computed for the remaining photographs of each *S*'s pupil for each slide.

Stimulus Materials

Twenty-four slides were prepared in Hess' laboratory, from magazine pictures selected by the present investigators. Five experimental slides were pictures of attractive women, nude or seminude, from *Playboy* or *Esquire* magazines. Another five experimental slides were pictures of attractive men. Two of these were from physique magazines. These pictures of men, unlike those of the women, were "proper" enough to appear in a family oriented evening newspaper.

There were also 12 control slides. Six of these each contained five numbers, one in each of the four corners and one in the middle (as previously described by Hess, 1965), and 6 were photographs of landscapes, taken from *National Geographic* and *Life* magazines. Hess' group has used both number slides and pictorial slides as control stimuli. The landscapes were used in the present study because they, like the experimental slides, were photographs of objects with greater detail than the number slides. (There were also two experimental slides showing bloody mutilated bodies. These slides were not germane to

the concerns of the present paper, but were instead inserted into this study to provide evidence on the unrelated issue of whether pupillary constrictions or dilation occurs in response to repugnant stimuli. Data concerning them will not be reported here.) The control slides were alternated in their presentation with the experimental slides. Each experimental slide was preceded by either a number slide or a landscape slide. Both the female and male slides were preceded and followed as often by number slides as by landscape slides. Six different orders of slides were used so as to equate within the group of *Ss* as a whole the ordinal position of each type of slide.

Instructions and Procedure

Both *Es* recited the same instructions concerning the nature of the apparatus and the experiment, with the exception of a single line specified below. The *Ss* who came to the experiment wearing glasses other than contact lenses were asked to remove them. The *Ss* were told that the experiment was concerned with how a person looks at a picture. Pupillary response was not mentioned either during the experiment or afterwards, and no *S* indicated any suspicion that this was the topic of investigation. The *E* pointed out the screen and the lights in the box and told *S*, "Your task will be to look at each picture. Continue to look at each picture until the next one appears." The lights were turned off, and a number slide appeared on the screen. The *S* was asked to read the numbers, each of which was 1½ in. high on the screen 40 in. from his eyes. Three prospective *Ss* who were unable to read the numbers, were excluded from the experiment. The *Ss* were then shown three additional buffer slides each for 10 sec. One slide showed a baby, a second a girl on a boat, and a third was another number slide. The *E* then stopped the apparatus and checked the focus of his camera, and reminded *S* to keep his head as still as possible.

*E*₂, but not *E*₁, said at this point, in a "suggestive" tone of voice, "I think you will find some of the pictures interesting." This single line was the only difference between *Es* programmed into the script. This line was inserted for *E*₂ on the assumption that it was consistent with his other less explicit communications to *S* concerning his attitudes toward the study and toward *S*. Neither *E* spoke while *Ss* viewed the slides.

Subjects

The *Ss* were 51 male introductory psychology students at the University of Wisconsin. Four *Ss* were dropped because of poor quality of the photographs of their pupils. This resulted in a final group of 22 *Ss* for *E*₁ and 25 for *E*₂.

Like Hess and Poit (1960) and Nunnally et al. (1967) the present investigators did not scrutinize the heterosexuality of their *Ss*, but instead relied on the presumption that the majority of male college students have heterosexual interests.

RESULTS

For both *Es*' *Ss*, the mean pupil sizes for the landscape and number slides were almost identical. Therefore, these two types of control slides were combined in order to provide a stable base line with which to compare pupil size in response to the male and female slides.

Most recent investigators have analyzed pupillary response in terms of percentage change in pupil size. This mode of analysis seems inappropriate. The correlation, for both groups combined, between size of pupil diameter on the control slides and dilation to female slides (pupil size to female slides minus pupil size to control slides) was $-.28$ ($p = .05$). This means that the larger the pupil, the smaller the dilation, a finding closely analogous to the Law of Initial Values. This law, that the higher the basal level of autonomic activity the smaller will be the upward reactivity to a stimulus, is commonly found to hold for other psychophysiological measures, but has not received attention in recent reports on the pupillary response. An analysis of results expressing dilation as percentage change from pupil size for control slides would further depress differences that are already depressed by their relationship to a large initial pupil size. Mean percentage changes will be reported here in the interests of comparability with other research reports, but the significance tests were computed using values for pupil size adjusted for initial level. Percentage change was computed by dividing the difference between the pupil sizes for the control slides and each set of experimental slides by the pupil size for the control slides.

For both *Es*, *Ss*' pupils dilated to both the male and female slides. All of *E*₁'s *Ss* dilated to both female and male slides; however three of *E*₂'s *Ss* failed to dilate to the male slides, and two of these three *Ss* failed to dilate to the female slides. Mean dilation for *E*₁'s *Ss* was 3.2% to the female slides and 2.9% to the male slides. Mean dilation for *E*₂'s *Ss* was 3.8% to the female slides and 2.7% to the male slides.

Of central interest to this study is the comparative dilation obtained by each *E* for male and female slides. A regression model was used, as suggested by Benjamin (1963), to

adjust each individual's scores by the amount attributable to the correlation of the pupil size for experimental slides with pupil size for control slides. The formula used is a modification of the familiar formula for predicting scores on Y using the coefficient of regression of Y on X : $\hat{Y} = Y - b_{yx}(X - \bar{X})$.

The Y is the pupil size for experimental slides, X the pupil size for control slides, and \hat{Y} the adjusted pupil size for the experimental slides. These adjustments were performed separately for each S 's response to male and female slides, using regression coefficients computed for each E . The regression coefficients for E_1 were .945 for the female slides and .951 for the male. Comparable values for E_2 were .923 and .965.

Fourteen of E_1 's 22 S s showed larger pupil size while viewing female slides than male. The difference for the group as a whole was not significant, as shown by a direct difference t test using the adjusted scores ($t = .92$). In striking contrast, 20 of E_2 's 25 S s showed a larger pupil size viewing female slides than male, and the difference for the group was significant ($t = 2.54$, $p < .01$) as indicated by a one-tailed direct difference t test. The difference between the difference scores of the two E s' groups fell short of significance ($t = 1.53$, $p = .06$).^a

DISCUSSION

Although both E s found pupillary response to both male and female pictures, they found quite different results in the comparative dilation. If the two E s had been working in different laboratories, they would have reported contradictory findings. This result is consistent with the hypothesis that differing modes of interacting with S produce differing pupillary responses. These findings do not, however, identify the specific cues by which the two E s influenced their S s' responses.

The finding of pupillary dilation by 44 out of 47 male S s in response to male slides ap-

pears divergent from that of Hess et al. (1965) who found dilation for only two of five heterosexual S s. Hess et al., as previously mentioned, suggested that a differential response to the male and female slides may be a useful index of heterosexuality. Nevertheless, the present finding of dilation by male S s to male slides may be consistent with Hess' broader contention that pupillary response measures interest and positive evaluation. Heterosexual men in our society do many things that may be interpreted as reflecting both a strong interest in and a positive evaluation of other masculine men. An outstanding example is their lionization of athletes. Advertising agencies apparently exploit this interest in other men by using very masculine appearing men as models in their advertisements. Not all interest in other people need be erotic!

In evaluating Hess et al.'s results one must bear in mind that their heterosexual S s worked in Hess' laboratory and presumably were familiar with the hypotheses being tested. Any heterosexual young man in this circumstance would probably feel constrained to demonstrate his manhood by responding with lively interest to the female slides, but not to the male slides. The S s knew that E expected it of them, and they surely also expected it of themselves. In contrast, Hess et al.'s, homosexual S s had admitted their homosexual behavior to one of the investigators, and probably, therefore, felt under less constraint to demonstrate heterosexual interests.

Hess and his co-workers assume that the pupillary response is entirely beyond voluntary control. If the response can be altered by E 's behavior, as in the present study, it can probably also be influenced by self-instruction to attend with interest, or by self-expectations. Pupillary response is not isolated from the rest of S 's behavior, and apparently reflects, among other things, the general level of activation of the organism (Nunnally et al., 1967). Activation level is affected by many intrapsychic and interpersonal variables. In future research on pupillary response, investigators should control such variables as the personality of E , his interaction with S , the social atmosphere of the experimental sessions,

^a When the comparable significance tests were computed without the correction for initial level, the difference between pupil size for male and female slides was significant for E_2 ($t = 2.37$, $p < .02$) but not for E_1 ($t = .92$, $p = .36$), and the difference between the difference scores of the two E s was not significant ($t = 1.43$, $p = .15$).

and S's knowledge and expectations of the study.

REFERENCES

- BENJAMIN, L. S. Statistical treatment of the Law of Initial Values (LIV) in autonomic research: A review and recommendation. *Psychosomatic Medicine*, 1963, 25, 556-566.
- HESS, E. H. Attitude and pupil size. *Scientific American*, 1965, 212, 46-54.
- HESS, E. H., & POLT, J. M. Pupil size as related to interest value of visual stimuli. *Science*, 1960, 132, 349-350.
- HESS, E. H., & POLT, J. M. Pupil size in relation to mental activity during simple problem-solving. *Science*, 1964, 143, 1190-1192.
- HESS, E. H., & POLT, J. M. Changes in pupil size as a measure of taste difference. *Perceptual and Motor Skills*, 1966, 23, 451-455.
- HESS, E. H., SELTZER, A. L., & SELLEN, J. M. Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology*, 1965, 70, 165-168.

- HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "emotionality and perceptual defense." *Psychological Review*, 1950, 57, 229-234.
- LOEWENFELD, I. E. Mechanisms of reflex dilation of the pupil: Historical review and experimental analyses. *Documenta Ophthalmologica*, 1958, 12, 185-448.
- LOEWENFELD, I. E. Pupil size. *Survey of Ophthalmology*, 1966, 11, 291-294.
- MASLING, J. M. The influence of situational and interpersonal variables in projective testing. *Psychological Bulletin*, 1960, 57, 65-85.
- MCGINNIES, E. Emotionality and perceptual defense. *Psychological Review*, 1949, 56, 244-251.
- NUNNALLY, J. C., KNOTT, P. D., DUCHNOWSKI, A., & PARKER, R. Pupillary response as a general measure of activation. *Perception and Psychophysics*, 1967, 2, 149-155.
- ROSENTHAL, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.

(Received March 17, 1968)

(Continued from page 387)

- Multivariate Classification of Alcoholics by Means of the Minnesota Multiphasic Personality Inventory: Steven G. Goldstein* and Jamer D. Linden: Department of Medical Psychology, University of Oregon Medical School, 3181 South West Sam Jackson Park Road, Portland, Oregon.
- Clinician as an Aberrant Actuary: Illusory Correlation and the Incomplete Sentences Blank: B. James Starr* and Edward S. Katkin: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.
- Visual Perception of Sequential Numerosity by Normals and Retardates: Donald H. Thor* and Edward A. Holden, Jr.: Johnstone Training and Research Center, Bordentown, New Jersey 08505.
- Task Difficulty, Relaxation, and Anxiety Level during Verbal Learning and Recall: James H. Straughan* and W. Henry Dufort: Student Counseling Center, Western Washington State College, Bellingham, Washington 98225.
- Psychopathology and Semantic Interpretation of Ambiguous Words: Thomas B. Benjamin and Norman F. Watt*: Department of Social Relations, William James Hall, Harvard University, Cambridge, Massachusetts 02938.
- Sex and Dominance Cues in Good-Poor Premorbid Schizophrenia: Newton L. P. Jackson, Jr.*: Box A, Box 563, Camarillo State Hospital, Camarillo, California 93010.
- Preinstitutional Social Deprivation and IQ Changes among Institutionalized Retarded Children: Earl C. Butterfield* and Edward Zigler: Bureau of Child Research, University of Kansas, 3933 Eaton Street, Kansas City, Kansas 66103.
- Perceived Parental Behavior among Stutterers and Nonstutterers: Karen H. Bourdon and David E. Silber*: Department of Psychology, George Washington University, Washington, D. C. 20006.
- TAT Method for the Differentiation of Families of Schizophrenics, Delinquents, and "Normals": Marhta Werner, James R. Stabenau, and William Pollin*: Building 10, Room 2S239, Department of Health, Education, and Welfare, 9000 Rockville Pike, Bethesda, Maryland 20014.

* Asterisk indicates author for whom address is supplied.

(Continued on page 404)

SOCIAL COMPETENCE AND POSTHOSPITAL OUTCOME AMONG SCHIZOPHRENIC AND NONSCHIZOPHRENIC PSYCHIATRIC PATIENTS¹

BERNARD ROSEN,² DONALD F. KLEIN, SIDNEY LEVENSTEIN,

AND SIROON P. SHAHINIAN

Hillside Hospital, Glen Oaks, New York

Although Zigler and Phillips contend that an identical relationship between premorbid social competence and prognosis is found in both a schizophrenic and nonschizophrenic hospital population, supporting data are not presented. To test this hypothesis adequately the 3-yr. posthospital outcomes of 81 schizophrenic and 85 nonschizophrenic previously hospitalized patients were separately related to the Zigler and Phillips' Social Competence Scale. The results indicate that among schizophrenic patients social competence is positively related to two out of five indexes of posthospital outcome: incidence ($p < .001$) and frequency of rehospitalization ($p < .025$). Among nonschizophrenic patients, the relationship between incidence of rehospitalization and social competence is not statistically significant. With regard to frequency of rehospitalization a significant negative relationship ($p < .05$) was obtained. These results question the validity of the theoretical model of psychopathological development advocated by Zigler and Phillips.

Zigler and Phillips (1962) contend that psychopathology is a unitary phenomenon with the various diagnostic categories representing arbitrary segments derived from a maturity continuum. Central to their position is the hypothesis that the relationship between level of maturity and prognosis, generally accepted as existing among schizophrenic patients, is not unique to schizophrenia, but instead, cuts across all functional mental disorders.

Zigler and Phillips (1962) support their position by referring to the results of a previous study (Zigler & Phillips, 1961) which they contend indicates the existence of an identical relationship between level of premorbid maturity and prognosis in both a schizophrenic and nonschizophrenic hospital

population. Although the relationship of social competence to outcome is presented for a heterogeneous group of schizophrenic and nonschizophrenic patients, as well as for a subgroup of the schizophrenic patients in their sample,³ the relationship between premorbid social competence and posthospital outcome among nonschizophrenic patients is not reported.

The lack of specific analyses relating premorbid social competence to posthospital outcome for nonschizophrenic patients raises doubts as to the general prognostic efficacy of the Social Competence Scale as well as the validity of Zigler and Phillips' unitary theory of psychopathology.

¹ This study was supported in part by a grant provided by the Board of Directors, Hillside Hospital, and by Grants MH-10191 and MH-12273, and MH-08004 from the National Institute of Mental Health, United States Public Health Service. The authors would like to thank Robert E. Patton, Director of the Biostatistics Division of the New York State Department of Mental Hygiene, for his cooperation in making records available to us.

² Requests for reprints should be sent to Bernard Rosen, Hillside Hospital, 75-59 263rd Street, Glen Oaks, New York.

³ A reanalysis of the data presented in Study II (Zigler & Phillips, 1961, pp. 267-268) rejects the hypothesis that social competence is significantly related to posthospital outcome for either schizophrenic or nonschizophrenic patients. The recomputed probabilities using a Fisher's exact test are for nonschizophrenic patients .11, and for schizophrenic patients .098. (The probability value for schizophrenic patients is twice as large as that reported by Zigler and Phillips, 1961, p. 268; $p < .05$. This discrepancy may be due to an incorrect assumption made by Zigler and Phillips that the Fisher's exact test is tabulated as a two-tailed test).

METHOD

The sample studied consisted of 166 psychiatric patients described previously (Levenstein, & Pollack, 1966; Levenstein, Pollack, & Klein, 1966). Of the original sample of 176 patients 10 could not be followed-up and were excluded from this study.

Patients were diagnosed by the research psychiatrist according to criteria presented previously (Klein, 1967). The sample was diagnostically heterogeneous, 81 patients were diagnosed as schizophrenic; the remainder ($N=85$) consisted primarily of patients with affective and character disorders.

Follow-up interviews were conducted after the second and third year following discharge from Hillside Hospital, Glen Oaks, New York. Psychiatric rehospitalizations were further checked through the Biostatistics Division of the New York State Department of Mental Hygiene.

The social competence measure employed is the same as that specified by Zigler and Phillips (1961, 1962). The distribution of the social competence scores was divided as close to the median (1.17) of the entire sample as possible.

To assess posthospital outcome, the following measures, similar to those used by Zigler and Phillips (1961), were used: (a) incidence of psychiatric rehospitalization during the 3-yr. follow-up period;⁴ (b) the number of times the patient was rehospitalized during the follow-up period (frequency of rehospitalization); (c) the amount of time a rehospitalized patient had spent in the community between the time of discharge from Hillside Hospital and his first rehospitalization; (d) total length of psychiatric rehospitalization during the entire 3-yr. period; and (e) the average length of each rehospitalization.

RESULTS

The relationship between social competence and each outcome measure will first be presented for the entire sample and then separately for the schizophrenic and nonschizophrenic subgroups.

With regard to the incidence of rehospitalization among the entire sample, significantly more low-competence patients (54; 62.1%) were rehospitalized than high-competence patients (26; 32.9%; $p < .001$; one-tailed tests of significance were used where appropriate).

Table 1 contains the relationship between premorbid social competence and incidence of psychiatric rehospitalization for the two diagnostic groups. With regard to the schizophrenic subgroup, the results were consistent with those for the entire sample ($p < .001$).

⁴ Two patients who had never been rehospitalized, but who had committed suicide were included in the rehospitalized group for this variable only.

TABLE 1

INCIDENCE OF REHOSPITALIZATION AS A FUNCTION OF PREMORBD SOCIAL COMPETENCE

Competence	Schizophrenic			Nonschizophrenic		
	Hospitalized		Total	Hospitalized		Total
	N	%	N	N	%	N
Low	40	72.7	55	14	43.8	32
High	9	34.6	26	17	32.1	53

Note.—Total group: $\chi^2 = 14.10$, $df = 1$, $p < .001$, one-tailed; schizophrenic group: $\chi^2 = 10.73$, $df = 1$, $p < .001$, one-tailed; nonschizophrenic group: $\chi^2 = 1.17$, $df = 1$, $p = ns$, one-tailed.

However, among nonschizophrenic patients the incidence of rehospitalization was statistically similar for the two competence groups. The possibility exists, however, that although the incidence of rehospitalization for high and low-competence nonschizophrenic patients was similar, the nature of the rehospitalization experience differed. It can be hypothesized that among rehospitalized patients the frequency, total length, and average length of rehospitalization was greater, and that time from discharge to rehospitalization was shorter for low-competence patients, as compared with high-competence patients.

Table 2 examines the relationship between social competence and the aforementioned four measures of posthospital outcome for rehospitalized patients only. With regard to frequency of rehospitalization (Table 2), for the combined sample, the two competence groups could not be differentiated statistically. However, when the sample was divided into diagnostic subgroups, significant differences were observed. The rehospitalized low-competence schizophrenic patients were rehospitalized a significantly greater number of times than the rehospitalized high-competence schizophrenic patients ($p < .025$). Unexpectedly, among the rehospitalized nonschizophrenic patients the reverse was true ($p < .05$).

Table 2 contains the relationship between social competence and time between discharge and first rehospitalization. The two competence groups could not be differentiated statistically on this measure. This was true for the entire sample as well as for the two diagnostic groups.

TABLE 2

POSTHOSPITAL OUTCOME BY PREMORBID SOCIAL COMPETENCE FOR REHOSPITALIZED PATIENTS ONLY

Competence	Total		Schizophrenic		Nonschizophrenic	
	Low	High	Low	High	Low	High
<i>N</i>	53	26	39	9	14	16
Frequency of rehospitalization						
<i>M</i>	2.19	2.16	2.38	1.44	1.64	2.56
<i>SD</i>	1.13	1.29	1.17	0.68	0.81	1.37
<i>t</i>	0.11 ^a		2.28 ^{**a}		2.13 ^{*b}	
Time from discharge to first rehospitalization (in mo.)						
<i>M</i>	8.96	10.24	8.08	7.78	11.43	11.62
<i>SD</i>	9.46	10.04	8.72	10.00	10.88	9.79
<i>t</i>	0.57 ^a		0.09 ^a		0.05 ^a	
Total length of rehospitalization (in mo.)						
<i>M</i>	12.30	8.76	12.74	9.00	11.07	8.62
<i>SD</i>	9.14	8.75	7.99	10.04	11.68	7.93
<i>t</i>	1.70 ^{*a}		1.18 ^a		0.66 ^a	
Average length of each rehospitalization (in mo.)						
<i>M</i>	7.40	4.96	6.87	7.89	8.86	3.31
<i>SD</i>	8.43	6.98	6.91	10.41	11.52	2.75
<i>t</i>	1.24 ^a		0.51 ^a		1.80 ^c	

^a One-tailed test of significance.^b Two-tailed test of significance.^c Probability value corrected for unequal variances.^{*} $p < .05$.^{**} $p < .025$.

The total length of rehospitalization (Table 2) was significantly greater among low-competence patients when compared with the high-competence patients in the combined sample ($p < .05$), but there were no significant differences within either of the two diagnostic groups.

When the two competence groups were compared with regard to average length of each rehospitalization (Table 2), there was no significant difference between the two groups for the entire sample, or either of the two diagnostic subsamples.

DISCUSSION

The present study contradicts Zigler and Phillips (1962) as to the existence of an "identical relationship between pre-morbid social competence and prognosis in both a schizophrenic and non-schizophrenic hospital population [p. 216]." This contradiction can

be understood if one examines those significant findings common to the two studies, that is, the incidence of rehospitalization for the total group and the schizophrenic subsample. Here, the present study is in complete agreement with Zigler and Phillips (1961). In both studies a significant relationship was reported between social competence and incidence of rehospitalization for the total heterogeneous patient group, and for the schizophrenic subgroup. One might then erroneously infer that since the hypothesized relationship exists for the schizophrenic patients, and still holds when this group is combined with a nonschizophrenic group, that the relationship must also hold for nonschizophrenic patients. Methodologically, this study emphasizes the error involved in applying this type of reasoning to psychiatric research.

The results obtained with the variable

"frequency of rehospitalization," not included in Zigler and Phillips' (1961) study, suggests an inverse, rather than an "identical" relationship between premorbid social competence and posthospital outcome among schizophrenic and nonschizophrenic patients. The lack of a significant positive relationship between premorbid social competence and posthospital outcome for nonschizophrenic patients questions the theoretical model of psychopathological development advocated by Zigler and Phillips.

REFERENCES

KLEIN, D. F. Importance of psychiatric diagnosis in prediction of clinical drug effects. *Archives of General Psychiatry*, 1967, 16, 118-126.

LEVENSTEIN, S., POLLACK, M., & KLEIN, D. F. Follow-up study of formerly hospitalized psychiatric patients: Procedural considerations in data collection. *Journal of Hillside Hospital*, 1966, 15, 152-164.

LEVENSTEIN, S., KLEIN, D. F., & POLLACK, M. Follow-up study of formerly hospitalized voluntary psychiatric patients: The first two years. *American Journal of Psychiatry*, 1966, 122, 1102-1109.

ZIGLER, E., & PHILLIPS, L. Social competence and outcome in psychiatric disorder. *Journal of Abnormal and Social Psychology*, 1961, 63, 264-271.

ZIGLER, E., & PHILLIPS, L. Social competence and the processes-reactive distinction in psychopathology. *Journal of Abnormal and Social Psychology*, 1962, 65, 215-222.

(Received March 25, 1968)

(Continued from page 400)

- Phasic Electrodermal Orienting Response in Chronic Schizophrenics: II. Response to Auditory Signals of Varying Intensity: Alvin S. Bernstein*: Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, New York 11203.
- Anxiety Level of Schizophrenic and Control Patients and Their Parents: Amerigo Farina* and Jules D. Holzborg: Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.
- Affective Experience in Psychotherapy: Kenneth I. Howard,* Davic E. Orlinsky, and James A. Hill: Department of Psychology, Northwestern University, Evanston, Illinois 60201.
- Increases in Hypnotizability Resulting from a Prolonged Program for Enhancing Personal Growth: Charles T. Tart*: Department of Psychology, University of California, Davis, California 95616.
- Effects of a "Reformed Sinner" and a "Lapsed Saint": Thomas Harford and Leonard Solomon*: Department of Psychology, Boston University, 111 Cimmington Street, Boston, Massachusetts 02215.
- Determinants of Institutional Release and Prognosis in Mental Retardates: William Vogel,* Karoly J. Kun, and Edward Meshorer: Department of Psychology, Worcester State Hospital, Worcester 1, Massachusetts 01604.
- Effect of Hypnotic Age Regression on the Magnitude of the Ponzo and Poggendorff Illusions: Michael Parrish, Richard M. Lundy,* and Herschel W. Leibowitz: Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.
- Strategies within Hypnosis for Reducing Phobic Behavior: Suzanne Larsen Horowitz*: Child Psychiatry Clinic, Stanford University Hospital, Stanford, California.
- Changes in Mental Status of Elderly Patients in Age-Integrated and Age-Segregated Hospital Milieus: Boaz Kahana* and Eva Kahana: Department of Psychology, Washington University, St. Louis, Missouri 63130.
- Psychoendocrine Studies of Patients with Spinal Cord Lesions: James W. McDaniel* and Alan W. Sexton: Department of Physical Medicine and Rehabilitation, University of Colorado Medical Center, 4200 East Ninth Avenue, Denver, Colorado 80220.

* Asterisk indicates author for whom address is supplied.

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editors: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

ARTHUR C. CARR, *New York State Psychiatric Institute and Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

RICHARD L. BLUMENTHAL
*New York State Psychiatric Institute
and Columbia University*

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JOSEPH L. FLEISS
*New York State Department of Mental
Hygiene and Columbia University*

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELLINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

"frequency of rehospitalization," not included in Zigler and Phillips' (1961) study, suggests an inverse, rather than an "identical" relationship between premorbid social competence and posthospital outcome among schizophrenic and nonschizophrenic patients. The lack of a significant positive relationship between premorbid social competence and posthospital outcome for nonschizophrenic patients questions the theoretical model of psychopathological development advocated by Zigler and Phillips.

REFERENCES

KLEIN, D. F. Importance of psychiatric diagnosis in prediction of clinical drug effects. *Archives of General Psychiatry*, 1967, 16, 118-126.

LEVENSTEIN, S., POLLACK, M., & KLEIN, D. F. Follow-up study of formerly hospitalized psychiatric patients: Procedural considerations in data collection. *Journal of Hillside Hospital*, 1966, 15, 152-164.

LEVENSTEIN, S., KLEIN, D. F., & POLLACK, M. Follow-up study of formerly hospitalized voluntary psychiatric patients: The first two years. *American Journal of Psychiatry*, 1966, 122, 1102-1109.

ZIGLER, E., & PHILLIPS, L. Social competence and outcome in psychiatric disorder. *Journal of Abnormal and Social Psychology*, 1961, 63, 264-271.

ZIGLER, E., & PHILLIPS, L. Social competence and the processes-reactive distinction in psychopathology. *Journal of Abnormal and Social Psychology*, 1962, 65, 215-222.

(Received March 25, 1968)

(Continued from page 400)

- Phasic Electrodermal Orienting Response in Chronic Schizophrenics: II. Response to Auditory Signals of Varying Intensity: Alvin S. Bernstein*: Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, New York 11203.
- Anxiety Level of Schizophrenic and Control Patients and Their Parents: Amerigo Farina* and Jules D. Holzberg: Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.
- Affective Experience in Psychotherapy: Kenneth I. Howard,* Davic E. Orlinsky, and James A. Hill: Department of Psychology, Northwestern University, Evanston, Illinois 60201.
- Increases in Hypnotizability Resulting from a Prolonged Program for Enhancing Personal Growth: Charles T. Tart*: Department of Psychology, University of California, Davis, California 95616.
- Effects of a "Reformed Sinner" and a "Lapsed Saint": Thomas Harford and Leonard Solomon*: Department of Psychology, Boston University, 111 Cimmington Street, Boston, Massachusetts 02215.
- Determinants of Institutional Release and Prognosis in Mental Retardates: William Vogel,* Karoly J. Kun, and Edward Meshorer: Department of Psychology, Worcester State Hospital, Worcester 1, Massachusetts 01604.
- Effect of Hypnotic Age Regression on the Magnitude of the Ponzo and Poggendorff Illusions: Michael Parrish, Richard M. Lundy,* and Herschel W. Leibowitz: Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.
- Strategies within Hypnosis for Reducing Phobic Behavior: Suzanne Larsen Horowitz*: Child Psychiatry Clinic, Stanford University Hospital, Stanford, California.
- Changes in Mental Status of Elderly Patients in Age-Integrated and Age-Segregated Hospital Milieus: Boaz Kahana* and Eva Kahana: Department of Psychology, Washington University, St. Louis, Missouri 63130.
- Psychoendocrine Studies of Patients with Spinal Cord Lesions: James W. McDaniel* and Alan W. Sexton: Department of Physical Medicine and Rehabilitation, University of Colorado Medical Center, 4200 East Ninth Avenue, Denver, Colorado 80220.

* Asterisk indicates author for whom address is supplied.

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editors: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

ARTHUR C. CARR, *New York State Psychiatric Institute and Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

RICHARD L. BLUMENTHAL
*New York State Psychiatric Institute
and Columbia University*

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JOSEPH L. FLEISS
*New York State Department of Mental
Hygiene and Columbia University*

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt

Journal of Abnormal Psychology

Department of Psychology

N. Y. State Psychiatric Institute & Columbia University

722 West 168th St.

New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR	MYRTLE HENDRICKS	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Prince and Lemon Sts., Lancaster, Pa. 17604

and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

SOURCES OF VARIANCE IN RESPONSES TO QUESTIONNAIRES AND IN BEHAVIOR

RUDOLF H. MOOS¹

Stanford University School of Medicine

Sixteen psychiatric patients were observed in and were asked to describe their reactions to six psychiatric ward subsettings. The findings, which replicated and extended earlier results, indicated (a) that persons, settings, and Person \times Setting interactions generally accounted for statistically significant and important proportions of the total variance, both in responses to questionnaires and in actual behavior; and (b) that these proportions remained remarkably stable over a 3-mo. interval, except for a consistent tendency for the proportions attributable to settings to increase. Implications discussed include that different psychiatric ward subsettings may be differentially therapeutic to different patient groups, and that the findings bear on the upper limit of validity indexes obtainable with standard interview or rating predictions of patient behavior.

Several recent studies have demonstrated that persons, settings, and Person \times Setting interactions each account for statistically significant and important proportions of the total variance in responses to different questionnaires and in inferential ratings of behavior. The relative amounts of variance accounted for by different sources of variance vary importantly, depending upon the particular sample of persons and settings chosen for the study; however, it is generally true that the variance accounted for by the interaction between settings and persons is greater, sometimes substantially greater, than the variance accounted for by either persons or settings alone.

For example, Endler, Hunt, and Rosenstein (1962) and Endler and Hunt (1967) found that persons, settings, and Person \times Setting interactions each contributed significantly to behavioral variance in both anxiety and hos-

tility. Raush, Dittman, and Taylor (1959) and Raush, Farbman, and Llewellyn (1960), studying hyperaggressive and control children in a psychiatric ward milieu, found that interactions between child and setting were far more important in accounting for behavior than was either the child or the setting alone.

In a series of earlier studies (Moos, 1967, 1968b; Moos & Daniels, 1967), patient and staff responses to a number of different psychiatric ward subsettings, for example, individual therapy, group therapy, and community meeting, were analyzed for each of five different sets of adjectives reflecting trust, extroversion, security, involvement, and sociability. The results indicated that consistent individual differences among patients accounted for between 24 and 40% of the total variance; consistent setting differences always accounted for less than 5% of the total variance; and the Patients \times Settings interaction effect, with only one exception, accounted for a significant percentage of the total variance (between 17 and 26%). Results for staff were generally similar, except that consistent differences among individuals accounted for less variance than setting differences.

For example, one group of patients was more trusting in settings in which only one other person was present than in settings in which several others were present, whereas the results for a second group of patients were exactly the reverse. These findings need to be

¹The author is also at the Veterans' Administration Hospital, Palo Alto, California. The author wishes to express his appreciation to the patients and staff who participated in this research; to Eleanor Levine and Phyllis Nobel, who served as behavior observers; to William Lake, who provided valuable help in data analysis; to Bert Kopell and William Wittner, who importantly facilitated the work; and to Mark Abramson, John Adams, and Sheldon Starr, who permitted observations of their individual and group therapy sessions.

Requests for reprints should be sent to Rudolf H. Moos, Department of Psychiatry, Stanford University School of Medicine, Palo Alto, California 94304.

replicated using different samples for responses, persons, and settings. They are of particular importance for both the assessment and prediction of behavior and for the planning of maximally therapeutic environments, since they suggest that different types of settings may be maximally beneficial to different types of patients, and that psychological assessment procedures must systematically sample both persons and settings in order to predict behavior accurately.

The extent to which these results would be replicated if the data used were observations of actual behavior in different settings is important. Setting and Person \times Setting interactions might account for consistently greater or consistently less of the total variance when actual behavior is observed than when questionnaire responses are obtained. The extent to which the proportions of variance accounted for by different sources of variance change depending upon the particular time at which the sample is studied is also important. Does the percentage of variance attributable to persons or settings change with increasing familiarity with the environment and/or with successful psychiatric treatment?

PURPOSE

There were three major purposes for the present study: (a) to replicate earlier findings (Moos, 1968b) indicating that for psychiatric patients, differences among persons and Person \times Setting interactions account for much larger proportions of variance than setting differences; (b) to compare the percentages of variance accounted for by different sources of variance in questionnaire responses and in actual behavior; (c) to estimate the direction and extent to which the percentages of variance attributable to different sources of variance change with increasing familiarization with the environment and/or change in psychiatric status.

METHOD

Settings and Subjects

The research was conducted on an open-door, 30-patient psychiatric ward located in a large Veterans' Administration Hospital. This was a mixed ward with approximately 18 male and 12 female patients at any one time. The treatment was based upon

patient responsibility, decision making, and self-regulation within the framework of the ward community. There was a maximum emphasis on active coping with the milieu, on behavior change, and on the process of making practical plans for release, for example, obtaining a job, handling financial crises, and reconstituting family relationships. Open communication between staff and patients and active patient participation were strongly encouraged.

During the study, community meetings and small-group therapy were each held 2 days a week. Individual therapy was formally arranged for the majority of the patients; for some of these, it was centered around frequent but brief contacts. The remainder of the treatment program consisted of recreational activities and various hospital workshop and industrial work assignments.

The patients who were initially admitted to the ward essentially constituted a representative sample of male and female veterans with diagnoses of neuropsychiatric disorders; that is, there were no specific criteria which were applied for admission to the ward.

Each of 16 patients (8 male, 8 female) was selected on the basis of several criteria: (a) their conversation had to be easily understandable; (b) they had to be able to cooperate in wearing a wireless radio transmitter, and also to be willing to answer a questionnaire about their reactions in each setting; (c) they had to be on the ward long enough for systematic behavioral change to occur. The average age of the patients was 38.8 yr. ($SD = 7.1$), and their average education was 13.4 yr. ($SD = 1.1$). The median length of hospitalization was approximately 3 mo. and the number of previous hospitalizations ranged from none to five. Eight of the patients carried a primary diagnosis of neurosis; 7 carried a primary diagnosis of schizophrenia, and 1 carried a primary diagnosis of character disorder.

Procedure

The research was completed as a part of a larger project studying the behavior of psychiatric patients by the use of behavior observers and a miniature wireless radio transmitter. The plans for the use of the radio transmitter were discussed in both staff and patient-staff meetings, and the patients participated in putting up television antennae around the ward, including bedrooms, dayroom, visitors' room, dining room, etc.

The 16 Ss were each observed twice in each of six different ward subsettings. These subsettings were: (a) an intake meeting (or discharge-planning meeting), in which the entire ward staff observed the patient being interviewed by one of the resident psychiatrists; (b) an individual therapy session; (c) a group therapy session in which the patient was a member of a group of approximately eight patients; (d) a community meeting in which all patients and staff participated; (e) lunch time, which occurred in a dining room on the ward; and (f) free time, which was unscheduled time on the ward

during which the patient was free to engage in any activity he chose. These six settings constituted a representative sample of regularly scheduled ward activities in which all patients participated.

The initial set of observations occurred during the first and second weeks of hospitalization. A second set of observations was made on 12 of these patients approximately 3 mo. later, the week before the patient had improved enough to leave the hospital. Four of the patients could not be observed a second time, either because they had left the hospital against medical advice or because they had to be discharged without adequate prior warning.

Three sets of observations were obtained on each patient in each setting: (a) questionnaire responses relevant to how patients were feeling; (b) behavioral observations of what patients were actually doing; and (c) recordings of what patients were saying. Only the results of the first two sets of observations are discussed in this report.

Patients were given a questionnaire immediately after their participation in each of the six settings. This questionnaire consisted of nine dimensions measured by three items each, on each of which patients rated their experience on a 1-5 scale. Four affect dimensions were measured, for example, anxiety (uneasy, fearful, anxious), depression (sad, lonely, blue), vigor (lively, active, energetic), pleasantness (pleased, lighthearted, elated). Two of the dimensions measured the perceived worth (worthwhile, important to me, glad I went) and the perceived therapeutic benefit of the setting (made me feel better, helped me gain self-confidence, helped me get well). The other three dimensions measured affiliation (I worked closely with others, people were friendly to me, I was included), participation (I worked hard, I was active, I talked freely and openly), and leadership (I was needed, I played an important part, I helped decide what to do).

The items for the first four dimensions were taken partially from the Nowlis (1965) Mood Adjective Check List and partially from other items used in the earlier study (Moos, 1968b). The items for the other five dimensions were directly related to Barker and Gump's (1964) notions of the possible effects of different types of behavior settings.

Each of the patients was also actually observed in each setting. These behavior observations were made by two Os. The Ss knew that they would be observed by one of these Os during the time they had the microphone on. The Os, who had spent several months on the ward and were integrated with the staff and familiar to all the patients, had been trained to categorize and note patient behavior in terms of 16 relatively simple behavior categories, for example, smiling, talking, arm movement, etc., taken from those used by Zinner (1963). During the intake and community meeting settings, O sat in the same room as Ss and staff, and during the group therapy and individual therapy settings O sat behind a one-way vision screen. During the lunch and free-time settings O followed the patient around as unobtrusively as possible.

In order to estimate the effects of these observations, Ss were observed both with and without the microphone on in both community meeting and group therapy. Results indicated that the average effects of wearing and being observed with the wireless radio transmitter microphone were very small, and that the differences that did occur suggested a slight increase in purposeful and a slight decrease in purposeless activity while wearing the transmitter. These results, which are similar to those of other studies in showing that observational effects may be quite small, are more fully discussed in another paper (Moos, 1968a).

RESULTS

The reliability of the behavioral observations was established by having the two behavior Os observe and rate for 16 reliability sessions after a period of initial training. The task was to note the occurrence of each coded behavior during each minute of a 20-min. observation session. The reliability correlations were very acceptable, ranging from a low of .77 to a high of .99.

The data were analyzed by using an analysis of variance design. Separate analyses of variance were calculated for each of the nine questionnaire scales and each of eight behavior categories in each of the two sets of observations.² With this technique it is possible to specify, for each of the questionnaire

²The analysis of variance employed a model which assumed that each of the three sources of variance was sampled randomly, even though this was not strictly the case. The logic of this approach was discussed by Endler, Hunt, and Rosenstein (1962) and by Moos (1968b). The statistical significance of these results is considered to be less important than either the results relevant to the proportions of variance accounted for by different sources of variance or the reproducibility of the results. The analyses of variance calculations used for the *F* tests follow those discussed on pages 199-202 in Winer (1962). For example, in a three-way random model, in which the interactions cannot be pooled, quasi *F* ratios with estimated degrees of freedom for both numerator and denominator must be calculated. These analyses indicated that in the first set of observations, person effects were statistically significant ($p < .05$) for all questionnaire variables and all but one behavior variable (nodding yes); person-setting interactions were significant for all variables; and setting effects were significant for five of the questionnaire variables (anxiety, vigor, participation, affiliation, leadership) and for all but two (smiling, general movement) of the behavior variables. Significance levels were generally similar in the second set of observations.

TABLE 1
PERCENTAGES OF VARIANCE ACCOUNTED FOR BY DIFFERENT SOURCES OF
VARIANCE IN QUESTIONNAIRE SCALES—FIRST OBSERVATIONS

Scale	Source						
	A	P	S	A × I	A × S	P × S	A × P × S
Anxiety	5.6	25.4	7.4	8.3	1.4	20.8	31.0
Depression	0.0	26.7	0.0	4.0	0.5	29.2	39.0
Vigor	2.0	25.0	5.0	5.0	1.9	29.2	32.1
Pleasantness	12.8	12.8	1.5	6.7	3.8	15.0	47.4
Worth	0.0	21.0	0.0	0.9	3.5	24.6	51.7
Therapeutic benefit	0.3	13.0	3.0	4.1	0.9	37.9	40.8
Participation	0.0	5.3	10.5	19.4	3.7	19.7	41.6
Affiliation	0.0	18.0	4.6	15.8	2.2	9.2	51.0
Leadership	1.2	11.5	0.8	10.5	1.6	24.4	50.0

Note.—Abbreviated: A = adjectives, P = persons, S = settings.

and behavior dimensions, the proportion of the total variance which is accounted for by persons, by settings, and by their interaction.

The percentages of variance accounted for by each source of variance were calculated for these random-effects analysis of variance models using the rationale and equations given by Gleser, Cronbach, and Rajaratnam (1965) and Endler (1966). These authors have pointed out that the analysis of variance technique can be used to estimate the relative magnitude of each individual component of variance, expressed as a percentage of the sum of the different variance components. The general logic of this technique involves breaking the expected mean squares into their various variance components and solving separately for each component.

Tables 1 and 2 give the percentages of variance accounted for by different sources

of variance in the questionnaire scales for the first and second sets of observations. Consistent person differences accounted for between 0 and 45%, consistent differences between settings for between 0 and 18%, and the interactions between persons and settings for between 9 and 38% of the total variance. Differences between adjectives, Adjective × Person, and Adjective × Setting interactions generally accounted for less than 10% of the total variance.

In both sets of observations the variance accounted for by Person × Setting interactions was greater than the variance accounted for by settings for all nine dimensions, and was approximately the same as the variance accounted for by persons. Also, the proportion of variance accounted for by persons was consistently greater than the proportion accounted for by settings. The amount of vari-

TABLE 2
PERCENTAGES OF VARIANCE ACCOUNTED FOR BY DIFFERENT SOURCES OF
VARIANCE IN QUESTIONNAIRE SCALES—SECOND OBSERVATIONS

Scale	Source						
	A	P	S	A × P	A × S	P × S	A × P × S
Anxiety	5.2	25.3	11.0	4.5	1.9	16.2	35.7
Depression	0.0	45.1	0.0	3.3	1.6	17.2	32.0
Vigor	7.3	10.6	7.3	4.1	3.3	34.9	32.5
Pleasantness	13.3	9.7	4.4	11.5	0.0	15.0	45.1
Worth	0.0	22.2	3.3	0.2	3.1	24.5	47.5
Therapeutic benefit	1.0	26.8	9.8	1.6	0.0	32.4	29.5
Participation	15.0	6.6	17.8	9.2	2.3	18.9	30.2
Affiliation	4.0	9.7	5.2	9.1	0.0	18.4	53.9
Leadership	4.2	0.1	9.9	17.2	2.2	23.3	43.1

Note.—Abbreviated: A = adjectives, P = persons, S = settings.

TABLE 3

PERCENTAGES OF VARIANCE ACCOUNTED FOR BY
DIFFERENT SOURCES OF VARIANCE IN BEHAVIOR
CATEGORIES—FIRST OBSERVATIONS

Category	Source			
	Persons	Settings	P × S	Within
Hand and arm movement	16.8	11.9	31.9	39.4
Foot and leg movement	27.4	10.0	26.7	35.9
Scratch, pick, rub	30.7	13.1	24.5	31.6
General movement and shifting	17.3	1.4	47.1	34.1
Nod yes	4.2	42.9	33.5	19.4
Smile	35.3	3.6	35.4	25.6
Talk	10.5	68.3	13.9	7.4
Smoke	41.9	7.1	20.7	30.2

ance accounted for by settings increased from the initial to the final set of observations for eight of the nine response dimensions, whereas the changes for person differences and Person × Setting interactions showed no consistent patterns.

Tables 3 and 4 show the percentages of variance accounted for by different sources of variance for the first and second set of behavior observations. The 16 behaviors had been classified into four categories on the basis of their intercorrelations, and the 8 behaviors shown in Table 3 represent the behaviors in each category which showed the most frequent occurrence. The behaviors of hand and arm movement; foot and leg movement; scratching, picking, or rubbing; nodding yes; and general movement and shifting fell into the general movement category; smiling fell into a category called expressiveness; and smoking and talking each fell into separate categories.

The percentage of variance accounted for by different sources of variance varied greatly depending upon the particular behavior being considered. Talking and nodding yes were more a function of the setting than of the person, although the Person × Setting interactions were substantial, especially for nodding yes. The other six behaviors showed generally similar results, in that the variance accounted for by persons and by Person × Setting interactions were both always greater than that accounted for by settings. The

variance accounted for by Person × Setting interactions generally tended to be as great as, or somewhat greater than, that accounted for by consistent person differences.

The findings indicate that there is much greater variation in the proportion of variance attributable to different sources of variance in actual behavior than in responses to questionnaires, especially in the proportion of variance attributable to settings, which varied from 1 to 68%. Most of the variance in talking was related to setting differences; that is, persons consistently tended to talk more in some settings than in others. On the other hand, most of the variance in smoking was related to person differences; that is, some people consistently smoked more than others. Furthermore, the amount of Person × Setting interaction variance was consistently high, indicating that, for example, some persons tended to smile more in Setting A than in Setting B, whereas other persons showed just the reverse tendency.

The changes from the initial to the final set of observations were very similar to those for the questionnaire responses. The variance attributable to settings showed an increase for seven of the eight behaviors (all except talking, which initially showed an extremely high setting variance), whereas the variance attributable to persons and to Person × Setting interactions showed no consistent changes.

Figures 1 and 2 show examples of interaction effects for the questionnaire variable of

TABLE 4

PERCENTAGES OF VARIANCE ACCOUNTED FOR BY
DIFFERENT SOURCES OF VARIANCE IN BEHAVIOR
CATEGORIES—SECOND OBSERVATIONS

Category	Source			
	Persons	Settings	P × S	Within
Hand and arm movement	17.2	13.8	29.6	39.3
Foot and leg movement	27.3	13.0	31.2	28.6
Scratch, pick, rub	26.3	18.2	27.6	27.9
General movement and shifting	23.1	4.4	48.2	24.3
Nod yes	4.6	56.5	21.3	18.5
Smile	33.4	8.3	36.1	22.3
Talk	7.4	60.1	19.9	12.5
Smoke	36.5	12.2	10.2	41.1

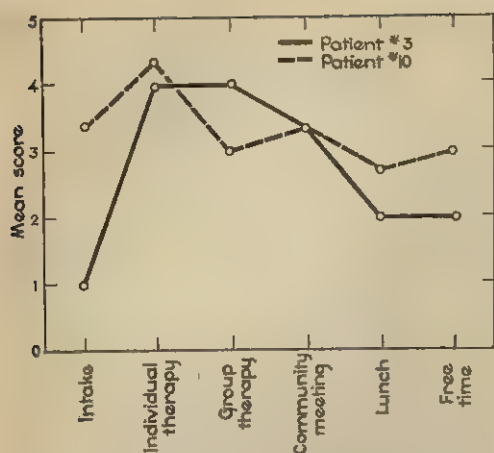


FIG. 1. Example of Person \times Setting interaction for the variable of therapeutic benefit of setting.

therapeutic benefit and the behavioral variable of the frequency of scratching and rubbing parts of the body. Figure 1 indicates that Patient 10 stated that he obtained more treatment benefit from intake meeting and less from group therapy than Patient 3, whereas there were no differences between these two patients in either individual therapy or community meeting. Figure 2 indicates that Patient 10 did more scratching and rubbing parts of his body in intake meeting and less in group therapy than Patient 3, whereas there were no differences between the two patients in individual therapy or community meeting. To the extent that scratching and rubbing parts of the body may be indicative of anxiety, an *O* would make very different estimates of the anxiety of these two patients, depending upon the setting in which they were observed.

DISCUSSION

These results essentially replicate and substantially extend those of previous studies. The questionnaire results for the two sets of observations were very similar to those obtained in an earlier study (Moos, 1968b), despite a different patient sample, the use of a different psychiatric ward, the sampling of different settings, and the measurement of different response dimensions.

The results also indicate that persons, settings, and Person \times Setting interactions

generally account for significant and important proportions of the total variance in behavior. These findings are consistent with those of Dittman (1963) and Ekman (1965), who have previously discussed the important influence of situational factors on nonverbal behavior, and Zinner (1963), who recorded observations of 56 specific behaviors on 90 airmen in basic training in 30 different real life settings and found that there were dependencies between settings and behaviors; for example, there were settings that either facilitated or inhibited smoking and talking to varying degrees.

The proportion of variance accounted for by different sources both in responses to questionnaires and in behavior were remarkably constant in the two sets of observations, except that the variance attributable to setting differences generally increased over time. These findings probably indicate that settings tend more consistently to elicit particular behaviors as persons become more familiar with the general milieu. As the 12 patients tested the second time had all improved in psychiatric condition, this change may be related to changes in psychiatric status. The current data do not allow a test between these alternative explanations, since psychiatric improvement and increasing familiarization with the milieu are confounded.

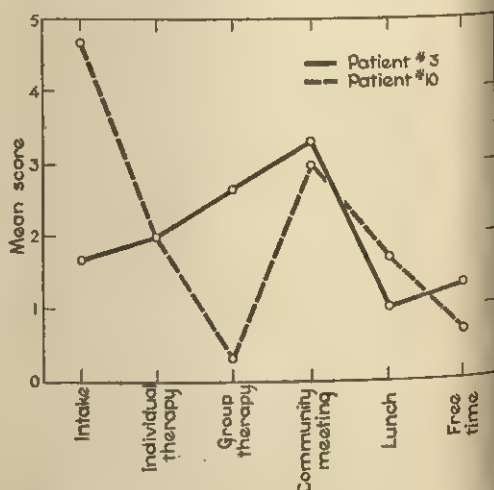


FIG. 2. Example of Person \times Setting interaction for the variable of scratching and rubbing parts of the body.

These findings replicate those of Raush et al. (1959, 1960) in indicating that situational factors (setting differences) may play a more potent role in behavior after treatment and/or experience with the general milieu. The two studies were so different (age of Ss, length of time between initial and final observations, range of settings, types of response dimensions studied, etc.) that these similarities indicate that the replicated findings may be highly general.

On the other hand, the specific sample of settings, persons, and response indicators used may affect importantly the relative proportions attributable to different sources of variance. For example, the results for the variable of anxiety were generally similar to those that Endler et al. (1962) obtained for their anxious questionnaires; however, the percentage of variance accounted for by modes of response (adjective differences) was less, whereas the percentages of variance accounted for by consistent individual differences were greater in this study. The former difference is probably due to the fact that Endler et al. sampled a wider range of anxiety indicators, whereas the latter difference may be due to the fact that psychiatric patients are generally high-anxiety individuals and, as such, would be expected to be more consistently anxious across settings than college student Ss.

Moos and Clesmes (1967) studied the amount of actual talking (by word count) of therapists in different individual therapy settings, and found that 18.5% of the variance was attributable to setting differences. This indicates that the variance related to settings is still statistically significant, even when all the settings sampled are similar; however, the actual proportion attributable to settings may decrease markedly when the range of settings is curtailed.

Further studies need to explore systematically these effects by experimentally or statistically (through Monte Carlo methods) sampling different populations of persons, settings, and response indicators. The degree of variation in the proportion of variance accounted for by different sources of variance clearly is much greater than has been assumed heretofore. This is particularly important since the percentage of variance accounted

for by consistent differences between individuals is related to the upper limit of the validity of predictions. If 9% of the total variance is accounted for by consistent individual differences, then the upper limit of validity is .3; if 16% of the variance is accounted for, then the upper limit is .4; and if 25% of the variance is accounted for then the validity may go as high as .5; etc. Thus, no predictions made from individual differences on the nine questionnaire response dimensions for the initial set of observations could have a validity better than .5. Predictions made from the behavior categories could be somewhat better, and might be as high as .6 for the category of smoking.

The general implications of these results for the measurement and prediction of behavior and for the planning of therapeutic milieus have been more fully discussed in an earlier paper (Moos, 1968b). It is perfectly clear that as long as there are large percentages of variance which are due to Person \times Setting interactions, the particular behavioral potential of the person-setting unit must be more fully studied. The most important aspects of this research area are that the percentages of variance accounted for by different sources of variance determine the extent of validity possible with the use of standard questionnaire and interview or rating measures of personality, and that different psychiatric ward sub-settings may be differentially beneficial to different patient groups.

REFERENCES

- BARKER, R. G., & GUMP, P. V. *Big school, small school*. Stanford: Stanford University Press, 1964.
- DETTMAN, A. T. Patterns of body movement. Paper presented at the meeting of the American Psychological Association, Philadelphia, September 1963.
- EKMAN, P. Communication through nonverbal behavior: A source of information about an interpersonal relationship. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition and personality*. New York: Springer, 1965.
- ENDLER, N. S. Estimating variance components from mean squares for random and mixed effects analysis of variance models. *Perceptual and Motor Skills*, 1966, 22, 559-570.
- ENDLER, N. S., HUNT, J. McV., & ROSENSTEIN, A. J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, 76(17, Whole No. 536).
- ENDLER, M. S., & HUNT, J. McV. S-R inventories of hostility and comparisons of the proportions of

- variance from persons, responses and situations for hostility and anxiousness. Unpublished manuscript, University of Illinois, 1967.
- GLESER, G. C., CRONBACH, L. J., & RAJARATNAM, N. Generalizability of scores influenced by multiple sources of variance. *Psychometrika*, 1965, 30, 395-418.
- MOOS, R. Differential effects of ward settings on psychiatric patients. *Journal of Nervous and Mental Disease*, 1967, 145, 272-283.
- MOOS, R. Behavioral effects of being observed: Reactions to a wireless radio transmitter. *Journal of Consulting and Clinical Psychology*, 1968, 32, 383-388. (a)
- MOOS, R. Situational analysis of a therapeutic community milieu. *Journal of Abnormal Psychology*, 1968, 73, 49-61. (b)
- MOOS, R., & CLEMES, S. Multivariate study of the patient-therapist system. *Journal of Consulting Psychology*, 1967, 31, 119-130.
- MOOS, R., & DANIELS, D. Differential effects of ward settings on psychiatric staff. *Archives of General Psychiatry*, 1967, 17, 75-83.
- NOWLIS, V. Research with the Mood Adjective Check List. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition and personality*. New York: Springer, 1965.
- RAUSE, H. L., DITTMAN, A. T., & TAYLOR, T. T. Person, setting and change in social interaction. *Human Relations*, 1959, 12, 361-378.
- RAUSE, H. L., FARBMAN, I., & LLEWELLYN, L. G. Person, setting and change in social interaction: II. A normal control study. *Human Relations*, 1960, 13, 305-332.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- ZINNER, L. The consistency of human behavior in various situations: A methodological application of functional ecological psychology. Unpublished doctoral dissertation, University of Houston, 1963.

(Received May 18, 1968)

A BRIEF COMMENTARY ON THE USEFULNESS OF STUDYING FEARS OF SNAKES

ALLAN COOPER,¹ JOSEPH B. FURST, AND WAGNER H. BRIDGER

Albert Einstein College of Medicine, New York

Therapy analogue studies using Ss with fears of snakes or rats have been used by several investigators to evaluate types of behavior therapy. The method had appeared to be an extremely convenient way of evaluating various therapies. In a small pilot study the authors were able to desensitize four Ss to fears of snakes or laboratory rats. This was consistent with reports in the literature. However, the two control Ss also overcame their fears of snakes. They had been instructed to hold the snake when it was presented to them and to act unafraid. They were also told that they would overcome their fears in this manner. The authors suggest that any "technique" that helps a person confront the fear of snakes may "cure" him. It is also suggested that studying the treatment of snake fears may be irrelevant to the understanding of treating clinical neuroses.

In this brief note, the authors suggest that evidence for the effectiveness of desensitization as a clinical technique should not be evaluated in studies of snake and rat phobias. The use of these animal studies initially appeared to be a brilliant stratagem. It provided a conveniently available phobic object that could be presented to a large pool of available Ss such as college students. The phobia was the same for all Ss. Change could be operationally defined in terms of distance from the snake that S could tolerate. However, despite all these apparent advantages, the usefulness of this type of study as an indicator of the effectiveness of desensitization in clinical practice was questioned. It is the authors' impression that both the type of S and the course of treatment are markedly different in college students who are fearful of snakes, and in people who are seen in actual clinical practice.

As part of a pilot study the authors successfully desensitized four students to their fear of harmless snakes or laboratory rats. These results replicated the findings of Lang, Lazovik, and Reynolds (1965) that desensitization can effectively cure a person's fear of snakes. Two other students served as control Ss. They were told that for the sake of science it was important to see if they could fool trained observers. They were instructed to act unafraid when a harmless snake was presented to them to hold. They were also told that by going through this procedure again and again, they would actually overcome their fears.

After an average of eight sessions the two control Ss were able to handle the snake and report that they were not afraid. Monitoring of their physiological reactions was in accord with their self-report.

Other investigators have reported results similar to those obtained with the control Ss in the present study. Strahley (1965) demonstrated that 16 Ss receiving "counterphobic" treatment were less afraid of a harmless snake than 16 Ss who received desensitization. Hogan and Kirchner (1967) report that significantly more of the 21 Ss treated with "implosion" were able to handle a laboratory rat after only one session than untreated control Ss.

One is impressed with all the various techniques that seem to be able to "cure" snake or rat phobias. However, the authors now suspect that the success of these techniques is due to "the natural fear extinction process" that one sees in human beings when they learn new skills. That is, when people learn to drive a car or dive into water, they soon overcome their fears with repeated exposure whether or not any "therapeutic" procedure is used to help them. It now appears that animal phobias in general are different from other clinical phobias. Marks and Gelder (1966) suggest that animal phobias have a much earlier onset, a more continuous course, come less frequently to psychotherapy, are associated with few or no other psychiatric symptoms, and respond better to behavior therapy. Moreover, the fear of snakes, in particular, is a fear that is not exposed to "the natural fear extinction process" since the person has little opportunity to encounter them in his daily life. In the case of fear of snakes, the therapeutic

¹Requests for reprints should be sent to Allan Cooper, Department of Psychiatry, Albert Einstein College of Medicine, Eastchester Road and Morris Park Avenue, Bronx, New York 10461.

aspects of various procedures may be irrelevant to the cure. The therapy may simply operate by providing repeated contact with the feared object.

The therapeutic procedure may also operate by providing explicit or implicit suggestion that the person will overcome his fears. Agras, Leitenberg, and Barlow (1968) report that Ss treated with desensitization but not given therapeutically oriented instructions or social reinforcement did not differ significantly from an untreated control group, while a group which did receive these variables was able to make progress in overcoming their fear of snakes. Thus, suggestion seems to play a role in the curative process.

Ordinarily, when one invokes the concept of "suggestion" it often appears as if what is meant is that S has simply been fooled or tricked into getting better. A more complex statement as applied to the fear of snakes and rats is that the belief that a person will get better may be sufficient motivation for him to face the fear and thereby expose the fear to the natural fear extinction process.

Studying the effectiveness of "treatments" for curing fears of snakes and rats may be irrelevant to a knowledge of the treatment of clinical phobias. Any "technique" that helps a person confront the fear of snakes may "cure" him. However, clinical experience suggests that this procedure of confrontation by itself will not be effective in the treatment of clinical phobias. It is also the authors' experience that desensitization is not nearly as effective with patients in clinical practice as it is with college students who are afraid of snakes.

The authors therefore question studying techniques for the treatment of fear of snakes, if this is being done for the purpose of understand-

ing the efficacy of these procedures in treating clinical phobias. It is suggested that despite the greater difficulty in procedures, only actual clinical phobias be studied to obtain data about techniques that are ultimately designed to treat clinical phobias.

SUMMARY

Therapy analogue studies, using Ss with fears of snakes or rats have been used by several investigators to evaluate types of behavior therapy. This technique has appeared to be an extremely convenient way of evaluating various therapies. However, as a result of curing two control Ss of their fear of snakes, the authors have come to the conclusion that studying the treatment of snake fears may be irrelevant to the understanding of treating clinical neuroses. Supporting evidence from other studies is cited.

REFERENCES

- AGRAS, S., LEITENBERG, H., & BARLOW, D. H. Social reinforcement in the modification of agoraphobia. *Archives of General Psychiatry*, 1968, 19, 423-427.
- HOGAN, R. A., & KIRCHNER, J. H. Preliminary report of the extinction of learned fears via short-term implosive therapy. *Journal of Abnormal Psychology*, 1967, 72, 106-109.
- LANG, P. J., LAZOVIC, A. D., & REYNOLDS, D. J. Desensitization, suggestibility and pseudotherapy. *Journal of Abnormal Psychology*, 1965, 70, 395-402.
- MARKS, I. M., & GELDER, M. G. Different ages of onset in varieties of phobias. *American Journal of Psychiatry*, 1966, 123, 218-221.
- STRAHLEY, D. F. Systematic desensitization and counterphobic treatment of an irrational fear of snakes. Unpublished doctoral dissertation, University of Tennessee, 1965.

(Received January 27, 1969)

PSYCHOLOGICAL DIFFERENTIATION AND PROCESS- REACTIVE SCHIZOPHRENIA¹

ROBERT CANCRO²

AND

A. ARTHUR SUGERMAN

Menninger Foundation, Topeka, Kansas

New Jersey Neuropsychiatric Institute, Trenton

1. Fifty-one consecutive male admissions to a psychiatric receiving hospital who had a staff diagnosis of schizophrenia were kept drug free for a period of testing. The rod-and-frame test (RFT) showed a curvilinear correlation coefficient of .28 with the Phillips Prognostic Rating Scale (PRS). The midrange on the RFT showed a good prognosis, with the process cases being at the extremes of the RFT scores. There was a significant relationship between RFT and duration of hospital stay at 1 and 2 yr. but not after 3 yr. of follow-up. The readmission pattern suggested that intermediate cases on the RFT leave the hospital sooner but return more frequently. The data failed to demonstrate a relationship between RFT and measures of conceptual plasticity, abstract attitude, severity of schizophrenic thought disorder, and severity of reality disruption.

The persistent problems in diagnosis of schizophrenia necessitated an improvement in classification. The concept of process-reactive schizophrenia was just such an effort to improve the nomenclature. The empirical finding that this distinction was useful has led to the attempt to clarify the underlying basis for the distinction. Many authors (Becker, 1956; Fine & Zimet, 1959; Herron, 1962; Kantor & Herron, 1965; Kantor & Winder, 1959; & Zimet & Fine, 1965) have suggested that classification is an expression of differences in the degree of psychological differentiation between schizophrenics. The process cases are viewed as less differentiated than the reactive cases.

There are no more reliable or better validated experimental measures of differentiation than those developed by Witkin, Dyk, Fater-son, Goodenough, and Karp (1962). These measures have been subjected to rigorous use in a theoretical framework of differentiation for a number of years. Two efforts (Bryant, 1961; Gibeau, 1965) to examine the process-reactive dimension in schizophrenia with these measures have produced slightly different

results. A major source of variance has been the use of drugged, as well as chronically institutionalized patients. Differences reported between process and reactive cases could easily be related to group differences in types, dosage, duration, or action of drugs. The effect of differences in the length of hospitalization between process and reactive patients could also account for some findings. Recognizing the essential need to control for the effects of drugs and duration of current hospital stay, the authors obtained a population that fulfilled these requirements. The attempts to demonstrate a linear relationship between position on the process-reactive continuum and the rod-and-frame test (Sugerman, 1962) and the sophistication of body concept test (Cancro, 1962) were unsuccessful.

An increasing number of reports by various investigators suggest that a linear relationship is not to be expected between these variables (Silverman, 1964; Voth, 1947; Witkin, 1965; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). These reports stimulated the present reexamination of this drug-free population for the presence of a non-linear relationship between psychological differentiation and the process-reactive classification. The secondary purpose was to seek out possible relationships between psychological differentiation, duration of hospitalization, and certain signs of the illness, specifically, ability to shift conceptual set, abstract

¹ Initial data collection was done at the State University of New York, Downstate Medical Center, with grant support from the Commonwealth Fund. The authors wish to thank Grant Dahlstrom and Herbert E. Spohn for their helpful comments and suggestions.

² Requests for reprints should be sent to Robert Cancro, The Menninger Foundation, Topeka, Kansas 66601.

attitude, severity of thought disorder, and degree of reality disruption.

METHOD

Subjects

Fifty-one consecutive male admissions to the wards of a large city psychiatric observation hospital, who had a senior staff diagnosis of schizophrenia, ranged in age from 18 to 44, and were fluent in English were admitted to this study.⁸ While the use of the staff diagnosis introduced an indeterminate amount of variability, it was chosen over the authors' diagnosis in order to avoid the problem of unconscious bias in the selection of cases. The age range was selected in order to screen out early adolescent reactions, involuntary states, and premature senility. The patients whose clinical picture was complicated by organic, neurological, or extraneous factors such as alcoholism, were rejected. An important methodological feature of this study was that patients received no tranquilizers or long-acting sedatives until after the period of testing and examinations, so that the clinical picture would not be contaminated by drug effect.

Procedure

Toward the end of the first week of hospitalization the patient received the rod-and-frame test (RFT, body erect), in Witkin's laboratory, without the technician knowing the patient's clinical history or mental status findings. The RFT score was used as the measure of psychological differentiation.

The patients were interviewed, and the material obtained was used to fill out the Prognostic Rating Scale (PRS, Phillips, 1953), which served to classify the patients on the process-reactive dimension. The patients were not divided into nominal categories on the basis of the PRS, but rather assigned their actual numerical score.

Patients were followed for a period of 3 yr., and the outcome measure was the total number of nights

⁸ The age range of the population was 18-44, $M=29.5$, $SD=7.4$. The mean number of years of education was 11.3, $SD=2.3$. The mean scaled WAIS vocabulary score was 9.3, $SD=3$. Twenty-eight patients (55%) had no prior psychiatric hospitalization. Of the 23 patients (45%) who were previously hospitalized, the mean duration of total prior hospitalization was 15.9 mo. with only 2 cases being hospitalized for a lifetime total of more than 36 mo. and no case for over 5 yr. Twelve patients with a history of previous hospitalization were classified as reactive so that patients with prior hospitalization were represented more highly in the reactive group (60%) than in the process group (40%). Thirty-four patients (66.7%) were in Hollingshead's Group V; 12 (23.5%) were in Group IV; 4 (7.8%) were in Group III; and only 1 patient was in Group I. There was no significant relationship between social class and the process-reactive classification in this sample.

spent in any mental institution during the specific time period under examination. The patients were sent to state and veterans' hospitals for treatment, without the treating hospitals being informed of their research status. There were no significant interinstitutional differences in duration of hospital stay for this sample.

The Metalog Test (Zubin & Windle, 1951) was used as the measure of ability to shift conceptual set and was scored blindly and independently. A Proverb List (Benjamin, 1944) was administered and independently scored for abstraction, using a scale developed by Meadow (Meadow, Greenblatt, Funkenstein, & Solomon, 1953). The severity of the thought disorder was clinically quantified by two separate and independently scored measures. The first of these was a rating of the severity of the formal signs on a 4-point scale. A score of 0 was given if no thought disorder was manifested during a 1-hr., mental status interview, done during the first few days of hospitalization. A rating of 1 was assigned if the patient showed circumstantiality, literalness, or concreteness. A rating of 2 was given if the patient showed autistic intrusions, predicative identifications, or loosening of associations. A rating of 3 was assigned if the patient showed perseverations, echolalia, blocking, neologisms, or incoherence. The patient was given the score of the most severe formal sign demonstrated. For example, if he showed loosening of associations and echolalia he was given a rating of 3. The second measure was the frequency of intrusions of formal signs of schizophrenic thought disorder in the proverb responses.

A clinical measure, designed to measure the degree of reality disruption on a 7-point ordinal scale of severity, was used (Cancro, 1962). Curvilinear correlation coefficients were used so that nonlinear relationships would be allowed to emerge. The RFT was treated as the nominal variable with the 51 cases divided into three equal groups of 17 each, representing the low, middle, and high range of differentiation. Pearson correlations were also calculated and F tests were done on the basis of the difference between the curvilinear and the Pearson correlations.

RESULTS

The RFT correlated with the PRS at .28, $F=4.0816$, $p<.05$. The intermediate group on the RFT consisted primarily of reactive cases as determined by the PRS. The most differentiated cases included most of the process patients and a very few reactives. The most undifferentiated group contained both process and reactive cases. In summary, the process cases tended to be at the RFT extremes, although more often at the highly differentiated rather than at the undifferentiated end, while the reactive cases were

mostly in the midrange and to a lesser degree in the undifferentiated end of the continuum.

The duration of hospital stay was correlated with RFT at four specific times during the 3-yr. follow-up period—6, 12, 24, and 36 mo. The RFT correlated with 6-mo. duration at .24, $p < .20$; with 12-mo. duration at .33, $p < .05$; with 24-mo. duration at .30, $p < .05$; and with 36-mo. duration at .22, $p < .20$.

At all four of these points it was the midrange group on the RFT that had the shortest duration of hospital stay. The least differentiated cases had an intermediate duration, and the most differentiated cases had the longest duration of hospital treatment. The susceptibility of the curvilinear correlation coefficient to variations in the manner of grouping patients into low, medium, and high RFT scores suggested the wisdom of grouping the patients by other criteria to confirm the relationship between RFT and outcome. Taking those 13 patients whose RFT score was within plus or minus .5 sigma from the mean as the midrange group produced a second grouping with 14 patients in the least differentiated and 24 patients in the most differentiated groups. This second grouping of the RFT correlated with 6-mo. duration at .19; with 12-mo. duration at .32, $p < .05$; with 24-mo. duration at .29, $p < .05$; and with 36-mo. duration at .23, $p < .20$. Witkin and associates (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954) reported in a study drawn from the same observation hospital that only 20% of their cases represented the midrange and that the remainder were equally split between high and low scores on the RFT. Grouping this patient population in a similar manner with the middle 20% ($n = 11$) representing the intermediate cases still gave essentially similar results. The third grouping of the RFT correlated with 6-mo. duration at .28, $p < .05$; with 12-mo. duration at .40, $p < .01$; with 24-mo. duration at .32, $p < .05$; and 36-mo. duration at .22, $p < .20$.

The Metalog Test correlated with the RFT at .32 but the curvilinearity was not significant. The Pearson correlation between the Metalog Test and the RFT was .27 with the most differentiated patients showing the most conceptual plasticity. There was a strong

TABLE 1
CURVILINEAR CORRELATIONS BETWEEN RFT
AND DURATION OF HOSPITAL STAY

Grouping	Mo.			
	6	12	24	36
First	.24	.33*	.30*	.22
Second	.19	.32*	.29*	.23
Third	.28*	.40**	.32*	.22

Note.— $N = 51$.

* $p < .05$.

** $p < .01$.

linear correlation between the Metalog Test and Wechsler Adult Intelligence Scale (WAIS) Vocabulary score of .46 in this sample. There was a marginal linear correlation between RFT and WAIS Vocabulary score and when the effect of the Vocabulary score was removed, the correlation between the RFT and the Metalog Test became non-significant.

There were no significant correlations between the RFT and the measures of abstract attitude, severity of thought disorder, and degree of reality disruption.

DISCUSSION

The importance of performing perceptual tests on a drug-free population cannot be overemphasized. Differences between patient groups, such as process and reactive, on a variety of psychological measures can be solely an expression of degree of tranquilization. It is also essential to control for length of hospitalization, especially since the process patients in any given study are likely to have been hospitalized longer than the reactive group to which they are compared.

The tendency to see process cases as sicker and more primitive than reactive cases, coupled with the tendency to see less differentiated people as more primitive and less stable, may have contributed to the establishment of a mental set that the relationship between these variables must be linear.

The finding that the midrange of differentiation, as measured on the RFT, consisted almost exclusively of reactive schizophrenics suggests that an intermediate degree of differentiation was correlated with good prognosis. The relationship between RFT and

duration of hospital stay tended to support the conclusion that differentiation contributes to the predictive power of the PRS. The cognitive stance and defensive preference of the extreme groups on the RFT differed sharply. Both extremes of differentiation were seen as overly committed to a particular ego style and lacking in the flexibility to shift according to the demands of the situation. From this point of view, patients in the midrange of differentiation would have a better prognosis, since they are less firmly committed to a particular style and can change more readily (Haronian & Sugerman, 1967). While this change may signal a true improvement it may only reflect the ability of the patient to comply with what the treating institution expects. One cannot help but be reminded of the warning about patients who improve rapidly only to lose their gains just as easily (Freud, 1964). This admonition, coupled with the failure of the RFT to significantly correlate with 6- and 36-mo. duration, suggested an examination of admission and discharge patterns over the 3 yr. of follow-up.

The curvilinear correlation between RFT and total number of admissions was .24, $p < .20$, while that between RFT and the total number of discharges was .23, $p < .20$. While not significant, the results were in the direction of the midrange on the RFT, having a higher number of readmissions, with the most differentiated cases having the fewest readmissions and the least differentiated cases an intermediate number. This was the same pattern that emerged with the discharges as well. The difference in duration of hospital stay between various RFT groups becomes nonsignificant because the higher readmission rate of the midrange and least differentiated patients begins to compensate for the longer initial hospital stay of the most differentiated patients at about 36 mo. This pattern of hospitalization suggests the long-term predictive elements of the PRS are not those that are related to psychological differentiation.

The failure of the RFT to correlate with any of the clinical variables studied suggests that these measures are not related to psychological differentiation. The restricted range of the Ss tested leaves open the possibility that

a relationship between various clinical signs and differentiation may exist in other types of patient samples. The clinical measures used in this study were affected to varying degrees by the severity of the schizophrenic process, and the failure to demonstrate a relationship to the RFT was consistent with the observation that psychological differentiation contributes to the form of the pathology and not its severity.

REFERENCES

- BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *Journal of Abnormal and Social Psychology*, 1956, 53, 229-236.
- BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Los Angeles: University of California, 1944.
- BRYANT, A. R. P. An investigation of process-reactive schizophrenia with relation to perception of visual space. Unpublished doctoral dissertation, University of Utah, 1961.
- CANCRO, R. *A comparison of process and reactive schizophrenia*. (Doctoral dissertation, State University of New York, Brooklyn) Ann Arbor, Mich.: University Microfilms, 1962. No. 5166.
- FINE, H. J., & ZIMET, C. N. Process-reactive schizophrenia and genetic levels of perception. *Journal of Abnormal and Social Psychology*, 1959, 59, 83-86.
- FREUD, S. Analysis terminable and interminable. In, *Complete psychological works of Sigmund Freud*. Vol. 23. London: Hogarth Press, 1964.
- GIBEAU, P. J. Field dependency and the process-reactive dimension in schizophrenia. Unpublished doctoral dissertation, Purdue University, 1965.
- HARONIAN, F., & SUGERMAN, A. A. Fixed and mobile field independence: Review of studies relevant to Werner's dimension. *Psychological Reports*, 1967, 21, 41-57.
- HERRON, W. G. The evidence for the unconscious. *Psychoanalytic Review*, 1962, 49, 70-92.
- KANTOR, R. E., & HERRON, W. G. Perceptual learning in the reactive-process schizophrenias. *Journal of Projective Techniques and Personality Assessment*, 1965, 29, 58-70.
- KANTOR, R. E., & WINDER, C. L. Process-reactive continuum: A theoretical proposal. *Journal of Nervous and Mental Disease*, 1959, 129, 429-434.
- MEADOW, A., GREENBLATT, M., FUNKENSTEIN, D., & SOLOMON, H. Relationship between capacity for abstraction in schizophrenia and physiologic response to autonomic drugs. *Journal of Nervous and Mental Disease*, 1953, 118, 332-338.

- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 352-379.
- SUGERMAN, A. A. Prognostic factors in schizophrenia: A developmental approach. Unpublished doctoral dissertation, State University of New York, Brooklyn, 1962.
- VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *American Journal of Psychiatry*, 1947, 103, 793-805.
- WITKIN, H. A. Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology*, 1965, 70, 317-336.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.
- ZIMET, C. N., & FINE, H. J. Primary and secondary process thinking in two types of schizophrenia. *Projective Techniques and Personality Assessment*, 1965, 29, 93-99.
- ZUBEN, J., & WINDLE, C. The prognostic value of the Metonym Test in a follow-up study of psychosurgery patients and their controls. *Journal of Clinical Psychology*, 1951, 7, 221-223.

(Early publication received February 14, 1969)

COMMENT ON "ARTIFICIAL INDUCTION OF POSTHYPNOTIC CONFLICT"

JOSEPH REYHER¹

Michigan State University

Sheehan's reason for discounting a previous investigation using posthypnotic conflict is factually incorrect, and his concept of "successful simulation" is challenged in this paper. His own procedures are questioned for several reasons: (a) he used a relatively weak posthypnotic conflict; (b) he did not incorporate any of the variables in his design which have been associated with the production of psychopathology in previous research; (c) he violated his own criteria for inducing posthypnotic conflict; and (d) his interpretation of his findings appears to be inconsistent. Recent investigations are cited which further document the pathogenic properties of the present author's procedures for inducing posthypnotic conflict.

In the course of developing a basis for his investigation of posthypnotic conflict, Sheehan (1969) discounted, on the basis of an improper control group, the results of investigations by Reyher (1961) and Moore (1964). Reyher reported the spontaneous repression of posthypnotically activated anger-aggression and the production of a wide range of psychopathology, including many psychosomatic symptoms. Using a modification of Reyher's procedures, Moore found that high achievers tend to repress the induced anger-aggression more than low achievers. He also reported the production of psychopathology, but the incidence was lower.

The control group described by Reyher (1961) was really a separate study in which the same Ss were given the experimental treatment under simulating and hypnotic conditions. The design was similar to one reported by Hilgard (1965, pp. 39-43) for sorting out order effects from main effects. The design was that highly susceptible Ss were given, in random or counterbalanced order, an experimental and a simulating condition. On one day S was hypnotized by E who left, and was awakened by a co-E who asked him to behave for E as if he were really hypnotized. On the other day, the co-E did not awaken S but merely gave a few simple suggestions that were executed; E returned and gave S the experimental treatment. The S was not advised as to

the second day until the completion of the condition on the first day. The main effects were determined through a comparison of control and experimental conditions occurring on the first day. In contrast to the high reliability (87%, $n = 23$) of producing symptomatic reactions with this procedure, the simulating condition produced no symptoms at all for five consecutive Ss; consequently, it was considered pointless to run more Ss (Interocular Traumatic Test, Edwards, Lindman, & Savage, 1963), even though not enough Ss had been run to sort out statistically the order effects from the main effects.

It was the order effect in combination with another point that led Sheehan to discount Reyher's and Moore's results. Sheehan (1969) contended:

The E, however, was unconcerned about motivating Ss to simulate successfully, and this procedure carried the strong implication that Ss should respond differently on the second day by changing their behavior from the responses they had shown previously [p. 16].

Although the word "unconcerned" does not reflect E's true attitude at the time, his use of the vague phrase "successful simulation" is not an improvement in design unless an operational definition or at least some sort of objective guidelines or criteria are provided to show how this is done by instructions to S. One of the best known facts of psychological research is the sensitivity of dependent variables to type of instruction. Quite aside from this lack of operational definition for

¹Requests for reprints should be sent to the author, Department of Psychology, Michigan State University, East Lansing, Michigan 48823.

successful simulation, the instruction to simulate already alters the demand characteristic of the research design (Reyher, 1967) for control Ss; additional motivating instructions further confound the issue. Even Orne (1959), who was a consultant to Sheehan's study, has recognized the intrinsic and considerable motivating properties of simulating instructions. The opportunity for a college freshman or sophomore to contribute to science and to put one over on the professor or *E* is quite sufficient, although as already noted, this alters the demand characteristics of the research for the simulators in unknown ways. Their unusual and high motivation probably accounts for why they overplay their roles (Reyher, 1968), scoring higher on the dependent variables than hypnotic Ss.

Sheehan's inclusion of Moore's study in his criticism of Reyher is incorrect, because Moore, unlike Reyher (1961), used separate groups of highly susceptible Ss for the simulating and experimental conditions. The control Ss were unknown to *E* and were asked to simulate by a co-*E*. Moore's results cannot be brushed aside so lightly.

No mention was made of Perkins' (1965) successful replication and extension of Reyher's procedures. He used the same control procedures as Moore and found that the experimental and simulating Ss were grossly disparate in number and kind of symptoms as well as GSR activity in response to the recognition of tachistoscopically presented conflict words. The GSR lends considerable objective support to the validity of the obtained psychopathology.

Sheehan's investigation is vitiated by his departure from Reyher's paradigm for inducing pathogenic posthypnotic conflicts, which he did not acknowledge, although its essential features were included in his own criteria. He used a reproachable act, which is a relatively weak drive or emotional state, instead of the sudden and intense posthypnotic activation of a basic drive. Consequently, his results were similar to the minor disturbances of other studies (Bobbitt, 1958; Huston, Shakow, & Erickson, 1934; Luria, 1932; Pruesse, 1967) which also used a "forgotten" reproachable act with posthypnotic suggestions that it will continue to

influence *S* in some way. He did not include any of those variables that put *S*'s defenses under great pressure and thereby produce many symptoms representing a wide range of psychopathology (Perkins, 1965; Reyher, 1958, 1961, 1967, 1968; Sommerschild, 1969). These variables are (a) a sudden surge of overwhelming rage toward an authority figure (b) coupled with an overwhelming impulse to destroy property belonging to said authority figure, and (c) activated suddenly by a posthypnotic cue (conflict word).

Modifications of the basic procedure substantially diminish its pathogenicity. Moore (1964) modified the procedure by reducing Variable 2 from overwhelming rage to nearly overwhelming and by eliminating Variable 3 entirely. Although he obtained a substantial amount of psychopathology, this was markedly less than what has been obtained by Perkins (1965), Reyher (1958, 1961, 1968), and Sommerschild (1969). Another investigation by Rokeach, Reyher, and Wiseman (1968) substitute reversals in beliefs for anger-aggression (Variables 1 and 2), resulting in even a lower incidence of symptoms. Veenstra (1969) eliminated Variable 2 and found that there was no spontaneous repression of the induced anger and, accordingly, there was no psychopathology.

The clinical relevance of Sheehan's data is weakened by his unintended departure from the proper paradigm (Reyher, 1962, 1963). Although he apparently endorses that part of the paradigm which stipulates that the hypnotically induced process must *spontaneously* produce *S*'s posthypnotic behavior, his induced paramnesia is incompatible with his intent because it contains specific posthypnotic suggestions, namely, you will feel "uneasy," "anxious," "terrible," "bad," and "worried." Aside from noting any differences between the experimental and simulating Ss, which there were, implications in regard to clinical phenomena cannot be drawn.

This author is puzzled by the way in which the differences between the experimental and simulating Ss were handled. On the one hand, Sheehan opined:

The pattern of results suggests that previous findings on the effect hypnosis has on personality after the

induction of conflict (e.g., Reyher, 1961) can be equally well explained by the demand characteristics of the experimental situation as by appeal to the effects of hypnotic treatment [p. 23].

On the other hand, he said of the hypnotic Ss:

The subjective reality of their feelings argues for changes and distortions in their behavior which accord with individual differences in the personality dynamics of Ss. In support of this, the evidence showed that the hypnosis had no single effect on Ss' personalities even though the same paramnesia was suggested to all Ss. Subjective reactions ranged from expression of somatic symptoms such as headaches to symptoms of unspecified distress, paranoid ideas and aggressive urges. Hypnotic Ss all reported that they experienced amnesia for the conflict material yet showed varying signs of breakdown of ego defenses. The most extreme reaction came from a hypnotic S who was agitated enough to hurl the TAT card at the E and say that he felt as if the police were watching him through a one-way screen which was placed in the room [p. 24].

These observations are *peculiar* to Sheehan's experimental Ss and are precisely the kinds of reactions which our (Moore, 1964; Perkins, 1965; Reyher, 1958, 1961, 1967, 1968; Sommerschild, 1969) experimental Ss have produced in great quantity, and which we consider to be of cardinal significance. Here is a method, finally, for spontaneously producing repression and psychopathology in the laboratory. We (Perkins, 1965; Reyher, 1958, 1961, 1967; Sommerschild, 1969) already have strong evidence from this line of research that the frequency of symptoms is an inverse function of the degree of repression and that the type of symptom is a function of degree of repression. Using an experimental design which included a replication group, Basch (1968) has verified the relationship between degree of repression and frequency of psychosomatic complaints using a physical health questionnaire and an inventory of items scored in terms of degree of drive representation.

REFERENCES

BASCH, J. Degree of repression and frequency of psychosomatic symptoms. Unpublished master's thesis, Michigan State University, 1968.
BOBBITT, R. A. The repression hypothesis studied in a situation of hypnotically induced conflict. *Journal of Abnormal and Social Psychology*, 1958, 204-212.

EDWARDS, W., LINDMAN, H., & SAVAGE, L. J. Bayesian statistical inference for statistical research. *Psychological Review*, 1963, 70, 193-242.
HILGARD, E. R. *Hypnotic susceptibility*. New York: Harcourt, Brace & World, 1965.
HUSTON, P. E., SEAKOW, D., & ERICKSON, M. H. A study of hypnotically induced complexes by means of the Luria technique. *Journal of General Psychology*, 1934, 11, 65-97.
LURIA, A. R. *The nature of human conflict*. New York: Liveright, 1932.
MOORE, W. F. The effects of posthypnotic stimulation of hostility upon motivation. *American Journal of Clinical Hypnosis*, 1964, 7, 130-135.
ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.
PERKINS, K. A. Repression, psychopathology and drive representation: An experimental hypnotic investigation of the management of impulse inhibition. Unpublished doctoral dissertation, Michigan State University, 1965.
PRUESSE, M. G. Repressors, sensitizers, and hypnotically induced hostility. Unpublished doctoral dissertation, University of Waterloo, 1967.
REYHER, J. Hypnotically induced conflict in relation to subception, repression, antisocial behavior and psychosomatic reactions. Unpublished doctoral dissertation, University of Illinois, 1958.
REYHER, J. Posthypnotic stimulation of hypnotically induced conflict in relation to psychosomatic reactions and psychopathology. *Psychosomatic Medicine*, 1961, 23, 384-391.
REYHER, J. A paradigm for determining the clinical relevance of hypnotically induced psychopathology. *Psychological Bulletin*, 1962, 59, 344-352.
REYHER, J. A reply to Levitt's comments. *Psychological Bulletin*, 1963, 60, 330-332.
REYHER, J. Hypnosis in research on psychopathology. In J. E. Gordon (Ed.), *Handbook of clinical and experimental hypnosis*. New York: Macmillan, 1967.
REYHER, J. *Hypnosis*. Dubuque, Ia.: Brown, 1968.
ROKEACH, M., REYHER, J., & WISEMAN, R. Importance of belief as a determinant of organization and change within belief systems. In M. Rokeach (Ed.), *Attitudes beliefs and values*. San Francisco, Calif.: Jossey-Bass, 1968.
SHEEHAN, P. W. Artificial induction of posthypnotic conflict. *Journal of Abnormal Psychology*, 1969, 74, 16-25.
SOMMERSCHILD, H. Posthypnotic conflict, repression and psychopathology. Unpublished doctoral dissertation, Michigan State University, 1969.
VEENSTRA, G. J. The effectiveness of posthypnotically-aroused anger in producing psychopathology. Unpublished master's thesis, Michigan State University, 1969.

(Early publication received April 4, 1969)

CONCEPT OF "PATHOGENESIS" IN PARENTS OF SCHIZOPHRENIC AND NORMAL CHILDREN¹

KEVIN M. MITCHELL²

Arkansas Rehabilitation Research and Training Center, Fayetteville

In this cross-validation study the Thematic Apperception Test was administered to five pairs of parents of normal and schizophrenic children. Parents by diagnostic category were matched on age, education, and social class. Each story was judged pathogenic, benign, or unscorable and a pathogenic score was tabulated for each *S* from the formula: pathogenic/pathogenic plus benign. It was found that mothers of the normal children told significantly fewer pathogenic stories than mothers of the schizophrenic children. Although fathers did not differ significantly, their mean pathogenic scores were distributed according to the hypotheses of the study.

Karon (1963) has conceptualized the schizophrenogenic mother as one who, when the needs of mother and child conflict, will indirectly satisfy her needs by manipulating her child in such a way that his behavior satisfies her needs. The covert nature of the mother's demands and her ambivalence lead to her child's inability to distinguish her motives from his and to seriously impaired ego functioning. Following this lead, Meyer and Karon (1967) and Mitchell (1968) found that mothers of schizophrenic children produced significantly more Thematic Apperception Test (TAT) stories than mothers of normal children in which the dominant individual either ignored expressed needs of the dependent individual, or took from the dependent individual when their needs conflicted. The authors termed these stories "pathogenic," and concluded that, regardless of the direction of causation, the concept of the schizophrenogenic mother is valid.

The purpose of the present study was to

cross-validate these findings on different samples of mothers of normal and schizophrenic children. Since available *Ss* included the fathers of the normal and schizophrenic children, a second hypothesis was that the fathers of the schizophrenic children would also tell significantly more pathogenic stories than the fathers of the normal children.

METHOD

The *Ss* were five sets of parents of normal and schizophrenic children who had been matched on age, education, and social class. Diagnosis of the schizophrenic children was based on interviews by two or more psychiatrists, as well as psychological tests. Normal children were nominated by the staff of a local junior college and a local high school.

Fathers were presented Cards 1, 2, 3BM, 6BM, 6GF, 7BM, 16, and 18BM. Mothers were presented Cards 1, 2, 3GF, 6BM, 6GF, 7GF, 16, and 18GF. Stories were typed on separate sheets of paper and were presented randomly to a clinical psychologist with approximately 4-yr. experience who blindly classified all stories as (a) pathogenic, (b) benign, or (c) unscorable. Classification was based on Karon's (1963) general formulation of "pathogenesis" which has been described in the preceding paragraph. However, if the dominant individual acknowledged and/or met the needs of the dependent individual, the story was scored benign (B). If there was no interaction, the story was judged unscorable (U).

In addition to the general criterion of meeting or not meeting the needs of a dependent individual, specific themes reflecting the pathogenic (P), B, and U categories had been collected from the two earlier studies (Meyer & Karon, 1967; Mitchell, 1968) and were used by both judges as guidelines in the present study. Half of the randomized stories were presented to a second judge, an advanced graduate student in clinical psychology, who also classified all the stories as P, B, or U. A pathogenic score was tabulated for each *S* using the formula:

¹ This study was supported, in part, through National Institute of Mental Health Grant 12306-02 and by a Research and Training Center Division Grant (RT-13) from the Social and Rehabilitation Service, Department of Health, Education, and Welfare, Washington, D. C.

² Also at the University of Arkansas. The author is grateful to Martha Werner, Section on Twin and Sibling Studies, Adult Psychiatry Branch, National Institute of Mental Health, and her colleagues for making their data available to him.

Requests for reprints should be sent to the author, Department of Psychology, Arkansas Rehabilitation Research and Training Center, West Annex, Fayetteville, Arkansas 72703.

$P/P+B$, where P is the number of stories judged pathogenic and B the number of stories judged benign.

RESULTS AND DISCUSSION

Interscorer reliability for the two judges as measured by a product-moment correlation was .79. Since both hypotheses were directional and differences in the opposite direction could not be explained within the framework of Karon's theory, all tests were one-tailed.

The t tests for matched pairs indicated that the mothers of the normal children told significantly fewer P stories than the mothers of the schizophrenic children ($t = 2.595$, $p < .05$). Mean P scores of the fathers of the normal and schizophrenic children were .43 and .61, respectively, but this difference was not significant ($t = 1.636$, $p < .10$).

The present data offer further cross-validation of Karon's concept of the schizophreno-

genic mother. Although the difference between the fathers was not significant, it was in the predicted direction. Significant differences might have been found between fathers if the N had been larger, or it may be that TAT themes specific to the "schizophrenogenic father" need to be identified.

REFERENCES

- KARON, B. P. The resolution of acute schizophrenic reactions: A contribution to the development of non-classical psychotherapeutic techniques. *Psychotherapy*, 1963, 1, 27-43.
- MEYER, R. G., & KARON, B. P. A study of the schizophrenogenic mother concept by means of the TAT. *Psychiatry*, 1967, 30, 173-179.
- MITCHELL, K. M. An analysis of the schizophrenogenic mother concept by means of the Thematic Apperception Test. *Journal of Abnormal Psychology*, 1968, 73, 571-574.

(Early publication received April 15, 1969)

PHYSIOLOGICAL EFFECTS OF RELAXATION TRAINING AND HYPNOTIC SUGGESTION

GORDON L. PAUL¹

University of Illinois

Comparative effects of hypnotic suggestion and brief relaxation training were evaluated with regard to reduction of subjective tension and distress (Anxiety Differential) and physiological arousal (heart rate, respiratory rate, tonic muscle tension, skin conductance). Three groups ($N=20$ each) of undergraduate female Ss participated individually for two sessions, 1 wk. apart, receiving (a) abbreviated progressive relaxation training as used in systematic desensitization therapy, (b) a hypnotic induction emphasizing direct suggestions of relaxation, heaviness, warmth, drowsiness, and sleep, or (c) a self-relaxation control procedure, included to evaluate the effects of merely resting quietly for an equal period of time with instructions to relax. In general, both relaxation training and hypnotic suggestion produced significantly greater effects than controls, and relaxation training resulted in significantly greater effects than hypnotic suggestion. The superiority of relaxation training was most pronounced in response systems not under direct voluntary control.

While it is assumed that multiple factors are involved in the application of systematic desensitization therapy (Paul, 1966, 1969a, 1969b; Lang, 1969) the "working hypothesis" has been that the major effective mechanism follows the counterconditioning paradigm which Wolpe (1958) termed the "reciprocal inhibition principle." According to this principle, the ability of given stimuli to elicit anxiety will be permanently weakened "If a response antagonistic to anxiety can be made to occur in the presence of anxiety-evoking stimuli so that it is accompanied by a complete or partial suppression of the anxiety responses . . . [Wolpe, 1958, p. 71]." Current research suggests that "anxiety" may best be thought of as a descriptive term for a complex pattern of response characterized by subjective feelings of apprehension and tension associated with

heightened physiological arousal, especially of the sympathetic branch of the autonomic nervous system (Spielberger, 1966). Thus, one basic requirement of the principle from which the systematic desensitization package was derived is that the *abbreviated* relaxation training used in systematic desensitization result in a reduced level of physiological arousal and subjective distress.

Even though controlled investigation has not found hypnotic suggestion to result in significant physiological change over that achieved by control procedures (Barber & Hahn, 1963), hypnotherapists make strong claims in this regard (e.g., Corley, 1964), to a point of attributing the success of systematic desensitization to "hypnosis" (see Cautela, 1966). The abbreviated relaxation training used with systematic desensitization does, in fact, share factors in common with hypnotic-induction procedures. Gill and Brenman (1959) noted that all standard-induction techniques (a) place extensive limits on sensory intake, (b) limit bodily activity, (c) restrict attention, (d) provide narrow and monotonous stimulation, and (e) alter the quality of bodily awareness. Barber and Calverley (1965) further noted that most hypnotic-induction procedures in the experimental literature include five components: (a) defining the situation as hypnosis, (b) instructing S to close his eyes, (c) administer-

¹ Appreciation is expressed to the Graduate College Research Board of the University of Illinois whose support made this study possible. Thanks are extended to Tom Brudenell, the polygraph operator, and to Ralph Trimble, E, for numerous hours devoted to pilot work and "debugging" of the apparatus in the months prior to the conduct of this investigation. At the time of this investigation, E had 2-yr. supervised clinical experience, including one semester's training in relaxation, desensitization, and hypnotic procedures under the investigator. Requests for reprints should be sent to the author, Psychological Clinic, University of Illinois, Champaign, Illinois 61820.

ing suggestions of relaxation, drowsiness, and sleep, (d) administering motivational instructions, and (e) suggesting that it is easy to respond to further suggestions. Cautela (1966) correctly pointed out that only two of the characteristics noted by Barber and Calverley are regularly included in brief relaxation training. However, all five elements mentioned by Gill and Brenman are also found in relaxation training. The major difference between relaxation training and hypnotic procedures appears to lie in the focus and effects involved in tension and release of gross muscle groups, defining the task as a passive "learning" situation in which *S* gains control rather than a "hypnotic" situation implying operator control, and the use of indirect suggestions of warmth, relaxation, etc., with instructions to maintain alertness rather than direct suggestions of relaxation, drowsiness, and sleep.

Two specific questions thus appear of major importance to the practice of systematic desensitization therapy, and to further understanding of the mechanisms involved in the effectiveness of the total treatment package: (a) Are reductions in physiological arousal and subjective distress producible within one to two sessions by *either* abbreviated progressive relaxation training or hypnotic suggestion? (b) Does progressive relaxation training produce effects which differ from hypnotic suggestion? Data relevant to these questions were obtained in the present study, where three groups of *Ss* participated individually for two sessions. The *Ss* of one experimental group received abbreviated relaxation training, while those of a second experimental group underwent a hypnotic induction focusing on suggestions of relaxation, drowsiness, and sleep. The *Ss* of the third group served as self-relaxation controls in order to assess the results of the two experimental procedures against the effects attributable to merely resting quietly for the same period of time with instructions to relax. The effects of the procedures were then evaluated on four physiological measures (forearm-muscle tension, heart rate, respiratory rate, skin conductance) and a self-report anxiety scale.

METHOD

Subjects

A total of 60 female *Ss* (17–23 yr. old) participated in the experiment as part of an introductory psychology research requirement. All *Ss* were screened for physical health and absence of drugs, but were otherwise unselected. No prior information was provided about the experiment, except that they were scheduled for two 1-hr. sessions 7 days apart, and that physiological responses would be monitored. The *Ss* were assigned to the three experimental conditions ($N = 20$ each) according to the time of arrival at the laboratory, following a repeating Latin square in blocks of three. The latter procedure, while maintaining random assignment, also counterbalanced for possible confounding due to time of day, season, or *E* fatigue.

Procedure

All *Ss* were run individually by one *E*, a 24-yr.-old married male advanced graduate student in clinical psychology, assisted by a polygraph operator—both under the regular supervision of the investigator. Upon entering the experimental room for the first session, *S* completed a set of paper-and-pencil scales, requiring a maximum of 10 min., the results of which will be reported elsewhere. She was then introduced into an electrically shielded enclosure, and placed in a semirecumbent position in a recliner chair. During the next 5 min., *E* and the polygraph operator attached a strain gage and electrodes for physiological recording, being careful to explain the purpose of each, and assuring *S* that no shock or other unpleasanties would be involved. The polygraph operator then retired to the adjoining control room while *S* completed a pretreatment anxiety scale—the Anxiety Differential (Husek & Alexander, 1963; Paul, 1966). The *E* then related the following general introduction to *Ss* in all conditions:

This experiment is being done in order to learn more about the clinical treatment of anxiety. Several of the procedures which we will use are taken directly from the clinical setting. Of special interest in the treatment of anxiety are the specific physiological manifestations of tension or physiological arousal. We suspect that normal subjects, such as yourself, with differing personality characteristics will show differing physiological systems involved in relaxation and stress, even of very mild sorts.

The tests you just took are designed to indicate your major personality characteristics; and if you're interested in their results you can contact us later in the semester and we'll be happy to go over them with you. We hope that this experiment will lead to a technique whereby psychologists can quickly localize the areas where a given person's tension and anxiety are expressed, and the best way of overcoming them.

Needless to say, we consider this work to be of utmost importance. Your cooperation will be neces-

sary in order to insure the validity of the results. As we indicated in your appointment letter, we'll have you come back again next week at the same time. We'll be asking you to take the Blue Test [Anxiety Differential] again at the end of today's session and at the beginning and end of the session next week, but you won't have to take any of the other tests again.

The specific introduction for each experimental condition (described in the following paragraph), followed without interruption. After questions were answered, a 10-min. silent adaptation period began, with Ss instructed to sit quietly with eyes open "while we finish calibrating the instruments." The last minute of this adaptation period was marked on the polygraph chart paper to serve as the pretreatment basal level (Period 1) for all physiological measures. All Ss were then instructed to close their eyes, and the next 2 min. were devoted to presentation of imaginal stimuli in conjunction with another study. The treatment or control procedures were then undertaken for 24-28 min. ($M = 27$ min.) per S, matched across conditions with the progressive relaxation Ss. Three 20-sec. time periods (Periods 2, 3, and 4) were marked on the chart paper of relaxation Ss, corresponding to $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$ through the training procedure, to provide scorable samples of physiological activity. Identical time periods for scoring were then determined for the hypnotic and control Ss within the same block. The last minute of recording after completion of training, but prior to termination instructions, provided the equated time interval for posttreatment assessment (Period 5). Immediately following termination instructions, a posttreatment Anxiety Differential was obtained.

On Ss arrival for the second session, she was immediately placed in the recliner chair, and strain gage and electrodes again attached. The second pretreatment Anxiety Differential was obtained, followed by a brief review of the introduction and query about the instructions of the previous session. As in the first session, a 10-min. adaptation period followed with the last minute again serving as a common time for assessment of pretreatment basal level (Period 1). All Ss were then instructed to close their eyes, and the same treatment or control procedures involved in the first session were undertaken for 22-25 min. ($M = 23$ min.), equated across groups. Periods 2, 3, 4, and 5 were marked on the chart paper as in the first session. After Period 5, 2 min. were devoted to presentation of imaginal stimuli in conjunction with another study, followed by an additional 5 min. of the treatment or control procedure. Termination instructions were then given, and a posttreatment Anxiety Differential again obtained before disengaging S from the apparatus.

Brief relaxation training. The progressive relaxation procedures were those previously used with systematic desensitization in controlled outcome studies of individual (Paul, 1966, 1967) and group (Paul, 1968; Paul & Shannon, 1966) treatment. Immediately after the general introduction, Ss assigned to this experi-

mental condition were presented with the following specific introduction:

The rest of the period will be spent training you to learn how to relax—probably to become more relaxed than you've ever been before. In order for you to learn this, we will use a clinical technique which was first developed by Dr. Edmund Jacobsen of the University of Chicago back in the 1930s. Basically, the procedure hinges on the fact that complete relaxation is the absence of tension. If you are completely relaxed, it is physiologically impossible to be tense or anxious. Since the modified Jacobsen technique can be learned readily, the real advantage is that you can use relaxation to counter any anxiety or tension you might feel in the future.

The way in which we'll do this is to have you systematically focus your attention on the various gross muscle groups throughout your body, first tensing each group for a few seconds—holding them long enough for you to identify exactly where you feel tension, and what it feels like. Then, when I say "relax," I want you to immediately let go—to stop tensing—and merely focus your attention on what those muscles feel like as relaxation takes place. By first tensing your muscles, the level of tension increases over your current operating level, such that when you release the muscles, the level of tension drops below the point where you started. Each time you tense and release a muscle group, the resulting level of tension becomes progressively lower and lower—to the point where no tension is present at all. By focusing your attention on this process as it takes place, with practice, you can eventually reach the point where just thinking or recalling the experience of release is sufficient to bring about relaxation itself.

The order in which muscle groups would be taken was then briefly described, demonstrating those exercises frequently found difficult.

It's important that you remember to focus your attention only on the specific muscle group we're working with at any given time. Especially so that you may learn the procedure, since I'll be asking you to practice relaxation before you come back next week. Actually it will be to your advantage to master this procedure, since, if used correctly, it can help a lot in reducing tension around exam times, give you a "second wind" when you're tired from studying, and even help you fall asleep at night when you may be tense or have many things on your mind.

Since I'll ask you not to talk during the training procedure, except to answer my questions with a simple yes or no, head nod, or "umhum," do you have any questions at this time?—All right, just sit quietly with your eyes open for a few minutes while we finish calibrating the instruments.

After completion of the intervening procedures described in the preceding paragraphs, progressive relaxation training was undertaken following the

operations detailed in Paul's (1966) experimental manual,² with *S* first instructed to focus her attention on the dominant hand and forearm, to tense and hold these muscles, and then to release the tension immediately upon *E*'s instruction to "relax." During the release period, instructions to focus attention with indirect suggestions of warmth, relaxation, etc., were quietly repeated, for example:

Relax—attending only to the relaxation process in your right hand and forearm—becoming more and more relaxed—warmer and heavier—nothing for you to do but notice the pleasant sensations in your right hand and forearm as they become more and more relaxed.

The same procedure of tension, release, and focus with indirect suggestions was systematically followed through the other gross muscle groups (dominant upper arm, nondominant hand and forearm, nondominant upper arm, forehead, eyes and nose, cheeks and mouth, neck and throat, chest, back, and respiratory muscles, abdomen, dominant upper leg, calf, foot, nondominant upper leg, calf, and foot). Through completion of the neck and throat muscles, a minimum of two tension-release cycles were instituted, moving on to the next muscle group as soon as the present group was reported to be as completely relaxed as the previous one. Throughout the remaining muscle groups, *S* was asked whether the current group was as completely relaxed as the previous one following a single tension-release cycle. In all instances, however, the next muscle group was not undertaken until the current group was reported to be as completely relaxed as the previous, or until four tension-release cycles had been attempted. After all muscle groups had been completed, *S* was instructed to simply attend to the relaxation taking place in her entire body (for an additional 30 sec.), and then asked if she felt any tension at all. If tension was reported, further tension-release cycles were instituted in the particular muscle group. After a report of "no tension anywhere," *S* was instructed to "simply enjoy the state of relaxation—becoming even more relaxed—just listening to the sound of your own breathing," for an additional minute.

Termination instructions were then given: "In a few moments I'm going to count backwards from four to one. On the count of four, start moving your legs; three, your hands and arms; two, your head; and one, sit up and open your eyes." The Anxiety Differential was obtained immediately following the count of one. The *S* was then asked about any difficulties during training and reassured that even better results would be obtained the following session. While *S* was being disengaged from the apparatus, *E* instructed her to practice relaxation twice per day between sessions, for about 15 min.; indicating that

² The timing in relaxation training involved 5-7-sec. tension periods with 20-30-sec. release periods (correction of a misprint in the 1966 manual, page 119, in which release time is incorrectly given as 10-20 sec.).

she could become even more skilled with practice.

For the second session, preliminary instructions consisted of a query regarding the success of practice during the previous week, clarification of any problems involved, and a brief statement that the procedures would be the same as the previous session, but were likely to take less time as *S* had some practice.

During both sessions, the timing followed that presented in the experimental manual (Paul, 1966, see Footnote 2), except for the introduction of recording periods to allow systematic sampling of the polygraph record, without the activity generated by the active tension of muscles. Thus, an additional 20 sec. of attention to relaxation, with indirect suggestions, was introduced three times during the training procedure, upon completion of the nondominant upper arm (Period 2), the neck and throat (Period 3), and the dominant foot (Period 4).

Hypnotic suggestion. The hypnotic-induction procedure consisted of a composite previously developed by the investigator to serve as an "all purpose" relaxation induction for initial training of clinical students (Paul, 1965). Immediately after the general introduction *S*s assigned to this experimental condition were presented with the following specific introduction:

The rest of the period will be spent training you to learn how to relax—probably to become more relaxed than you've ever been before. In order for you to learn this, we will use a clinical technique which was first developed by Dr. M. V. Kline in New York in the early 1950s. Basically, the procedure utilizes hypnotic imagery to induce a state of deep relaxation. Since you can readily learn parts of the technique, the real advantage is that you can use relaxation to counter any anxiety or tension you might feel in the future.

The way in which we'll do this is to have you systematically focus your attention on a visual image of yourself. As I describe various sensations for your image, you'll find that you actually respond in accordance with the suggestions as we gradually induce deep relaxation with hypnotic suggestion.

I am sure that you will find hypnosis a most interesting experience. Your ability to be hypnotized depends on your willingness to cooperate. It has nothing to do with your intelligence or your willpower—if you pay close attention to what I say, and follow what I tell you, you can easily fall into a hypnotic sleep and experience the interesting things I will tell you to experience. Hypnosis is nothing fearful or mysterious—your willingness to cooperate and your interest are what I ask for.

The *S* was then asked if she had been hypnotized before, and told that the use would probably be a bit different than other procedures she might be acquainted with. The order of events, from visual

imagery, eye fixation of the image, and numerical deepening procedures were then described briefly:

Of course we wouldn't ask you to do anything which would make you appear foolish, nor is there any danger of your revealing things about yourself. In fact, I'll ask you *not* to talk during the training period. You'll be completely aware of everything that's going on at all times, and you'll be able to remember everything we do. I'll simply be helping you to learn how to relax completely. Actually, it will be to your advantage . . . [etc., exactly as in relaxation training instructions].

The standardized hypnotic induction was then undertaken, after the intervening procedures described in the preceding paragraphs. The induction was patterned after Kline's (1953) visual imagery technique, with an eye fixation induction emphasizing suggestions of heaviness, drowsiness, sleep, and relaxation, following Weitzenhoffer (1957) and Barber and Hahn (1963), but directed to S's image of herself. Direct suggestions of heaviness, warmth, and relaxation were then focused on S's own specific muscle groups, followed by numerical deepening procedures, patterned after induction techniques used in previous research by London and Davison.⁸ A test challenge for arm immobilization from the Stanford Hypnotic Susceptibility Scale (Weitzenhoffer & Hilgard, 1959) was included, after Period 4, for comparative purposes. The entire procedure was administered in a lulling, soothing, "hypnotic tone," emphasizing focus on E's voice, and direct suggestions of drowsiness, sleep, warmth, comfort, heaviness, relaxation, slow and regular breathing.

The termination instructions and procedure were the same as those used for relaxation training, with the additional statement that "When I reach the count of one you will be wide awake. You will feel alert, and well rested, as if you had taken a short nap." The S was instructed to practice the imagery procedure twice per day between sessions, for about 15 min.; indicating that she could become even more skilled with practice. For the second session, preliminary instructions exactly paralleled those of the relaxation group, and the hypnotic induction followed the same procedures as the first session.

Self-relaxation control. Immediately after the general introduction, Ss assigned to the control condition were given the following specific introduction:

The rest of the period will be spent with you merely sitting quietly and relaxing while we record your physiological responses on the dynograph.

We're especially interested in the different rates at which people become relaxed in various physiological systems, and the way in which you relax yourself. I'll just sit in the room with you while you rest quietly with your eyes closed. The only thing you'll have to do for this session is to be

sure you don't fall asleep, however, I'll have to ask you not to talk during that period, so do you have any questions now?—All right, just sit quietly with your eyes open for a few minutes while we finish calibrating the instruments.

After the intervening procedures were completed, Ss in this condition were told, "Now just get completely comfortable and rest quietly with your eyes closed for the remainder of the time. Try to become completely relaxed, but be sure you don't go to sleep." Termination procedures were the same as those for the relaxation group, with the explanation given "so as not to disturb the electrodes." The control procedures for the second session were identical to those of the first session.

Apparatus. All treatment and control conditions were conducted in an air-conditioned experimental room below ground. The shielded enclosure was a room within a room, measuring $7 \times 5 \times 6\frac{1}{2}$ ft., made of completely connected copper screening attached to a 2×2 in. wooden frame, with an earth ground. The polygraph, also with an earth ground, was located in an adjoining room, separated from E and S by a one-way glass. All electrodes were fed through a junction box within the enclosure, with a common S ground from the dominant leg, through shielded cable to the control room.

Physiological measures were continuously recorded on a 4 channel Beckman Type RB dynograph, equipped with 474A amplifiers, on Offner No. V654 chart paper driven at 5mm/sec. All electrodes were cleaned and recoated, where appropriate, after each S was run. Commercial electrode cream (Grass EC2) was used for all connections. The skin area for electrode placements was first cleaned with alcohol, and then rubbed vigorously for 1 min. with a gauze pad coated with electrode cream. The electrodes for recording muscle tension and skin conductance were first filled with cream and then taped in place within Dr. Scholl's No. 312 foam-ease corn pads, also filled with cream. Electrodes for recording heart rate were filled with cream and taped directly to the skin area, as was the ground. All electrodes, except the ground, were then lightly wrapped with elastic bandage to prevent movement artifact.

Muscle tension was recorded from the dominant forearm in order to obtain maximal amplitude of tonic muscle action potentials. Silver cup electrodes (Grass E-1B) were attached, with the active electrode centered on the common extensor and the reference electrode on the flexor. The signal was fed through a Beckman 9852 EMG integrator. Muscle tension, in mean microvolts/second, was then determined from the average of the greatest amplitude of integrated muscle action potential within each second of each recording period (60 sec. for Periods 1 and 5; 20 sec. for Periods 2, 3, and 4), excluding heart-rate artifact.

Heart rate was recorded from Grass E5G electrodes placed over the radial artery of each wrist, fed through a Beckman 9857 cardi tachometer coupler.

⁸ P. London and G. C. Davidson, personal communication, 1964.

TABLE 1

OVERALL MEANS AND CORRELATIONS FOR INITIAL RESTING LEVEL, PERIOD 1,
AND FINAL PERIOD, PERIOD 5, ON ALL MEASURES

Period	Anxiety Differential		Heart rate		Respiratory rate		Muscle tension		Skin conductance	
	M	SD	M	SD	M	SD	M	SD	M	SD
Session I										
Period 1	62.90	12.5654	86.22	12.0930	17.08	3.6305	22.04	18.9379	3.246	.58022
Period 5	50.17	11.9464	81.17	11.4687	14.30	3.6622	19.66	17.4039	2.533	.57472
Pearson <i>r</i>	.1646*		.8882		.3750		.9743		.6981	
Session II										
Period 1	56.40	15.4176	84.83	11.6723	17.57	3.6677	24.26	18.7846	2.899	.84625
Period 5	47.05	11.5310	79.33	11.2652	15.58	4.0171	21.99	17.6015	2.098	.64234
Pearson <i>r</i>	.5052		.8700		.4598		.9657		.7535	

Note.—Heart rate in beats/minute; respiratory rate in cycles/minute; muscle tension in mean microvolts/second; skin conductance in micromhos; *N* = 60; all Periods 1-5 differences significant at *p* < .001.

* Nonsignificant correlation; however, part-score analysis included for consistency—identical results obtained with raw score analysis.

Heart rate, in beats/minute, was then determined by counting pulses from cardi tachometer recordings, or R-wave peaks from electrocardiogram (EKG) recordings, within each recording period with appropriate conversions for the 20-sec. periods.

Skin conductance measures were obtained through a Beckman 9892A skin resistance coupler from Grass E-1B silver-silver chloride electrodes placed on the dorsal and volar (just in front of the metatarsals) surface of the dominant foot. The lowest resistance in K ohms was then determined for each recording period, and converted to conductance in micromhos.

Respiratory rate was obtained from a 10-in. mercury strain gage attached to an adjustable elastic belt which was fitted around the chest at the bottom of the sternum. The strain gage activated a Parks 270 plethysmograph, which fed directly through a Beckman 9853 strain gage coupler. Within each recording period, data were first transcribed according to respiratory period (number of seconds to complete 10 respiratory cycles), and then converted to respiratory rate in cycles/minute.

All data reduction was carried out by two trained clerks, overlapping scoring on 20 occasions for each measure, with interscorer reliability exceeding .999 on all measures, and additional spot checks maintained. Any missing data, due to equipment artifact (less than two for any measure), were replaced by the unweighted mean for the treatment group. Conversions and data analyses were performed by the IBM-7090 computer of the University of Illinois Computer Science Laboratory.

RESULTS

Randomized groups analyses of variance performed on all Period 1 physiological measures and pretreatment scales obtained in the first session found no significant between-groups differences for any measure (all *p*'s >

.20), indicating that the assignment procedure had adequately randomized and equated groups prior to the introduction of treatments. Examination of response to the arm immobilization challenge for the hypnotic group found "pass" rates of 90% for the first session and 85% for the second session, as compared to 14% and 36% on standardization samples (Hilgard, 1967), indicating a high degree of effectiveness, by usual standards, for the induction procedure.

The unweighted means for initial resting levels and final period of each session for the total sample are presented in Table 1 for general orientation. Table 1 reveals a change in the "relaxed" direction in each session, which was found to be highly significant (all *p*'s < .001) for all measures by *t* test of the difference between correlated means. However, inspection of the average change from pretreatment resting levels during Session I and II for each group, presented in Figures 1 and 2, respectively, shows that the relaxation group was the only one to consistently produce such changes across measures and occasions.

While within-session trends are of interest, the most relevant data concern the comparative change between groups from Period 1 to Period 5 during each session. In order to rule out possible influence of basal levels on the magnitude of response, as is frequently found with physiological data, a base-free measure of change is desirable (Benjamin, 1967; Tucker, Damarin, & Messick, 1966). There-

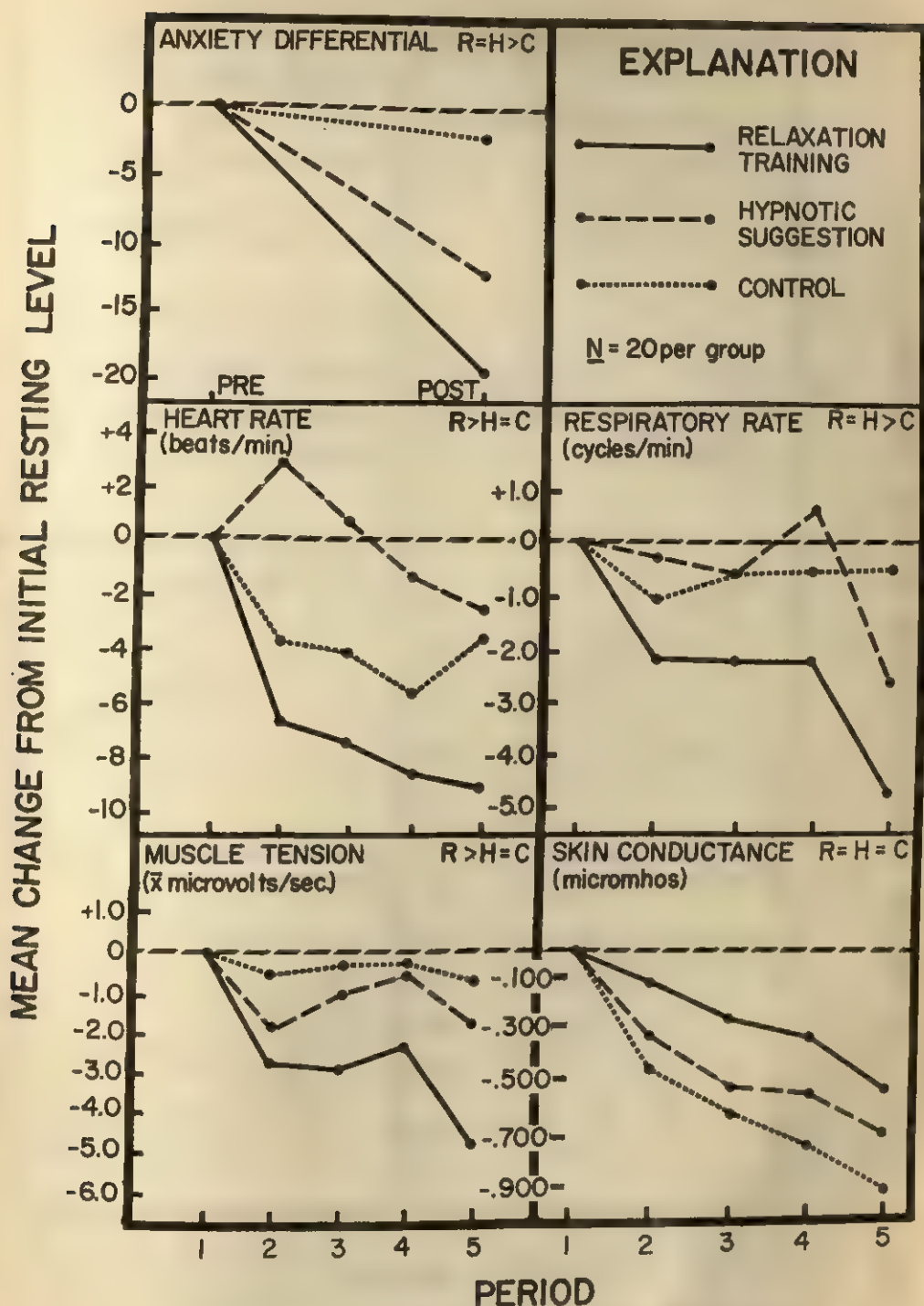


FIG. 1. Mean change from initial resting level on all measures during Session I.

MEAN CHANGE FROM INITIAL RESTING LEVEL

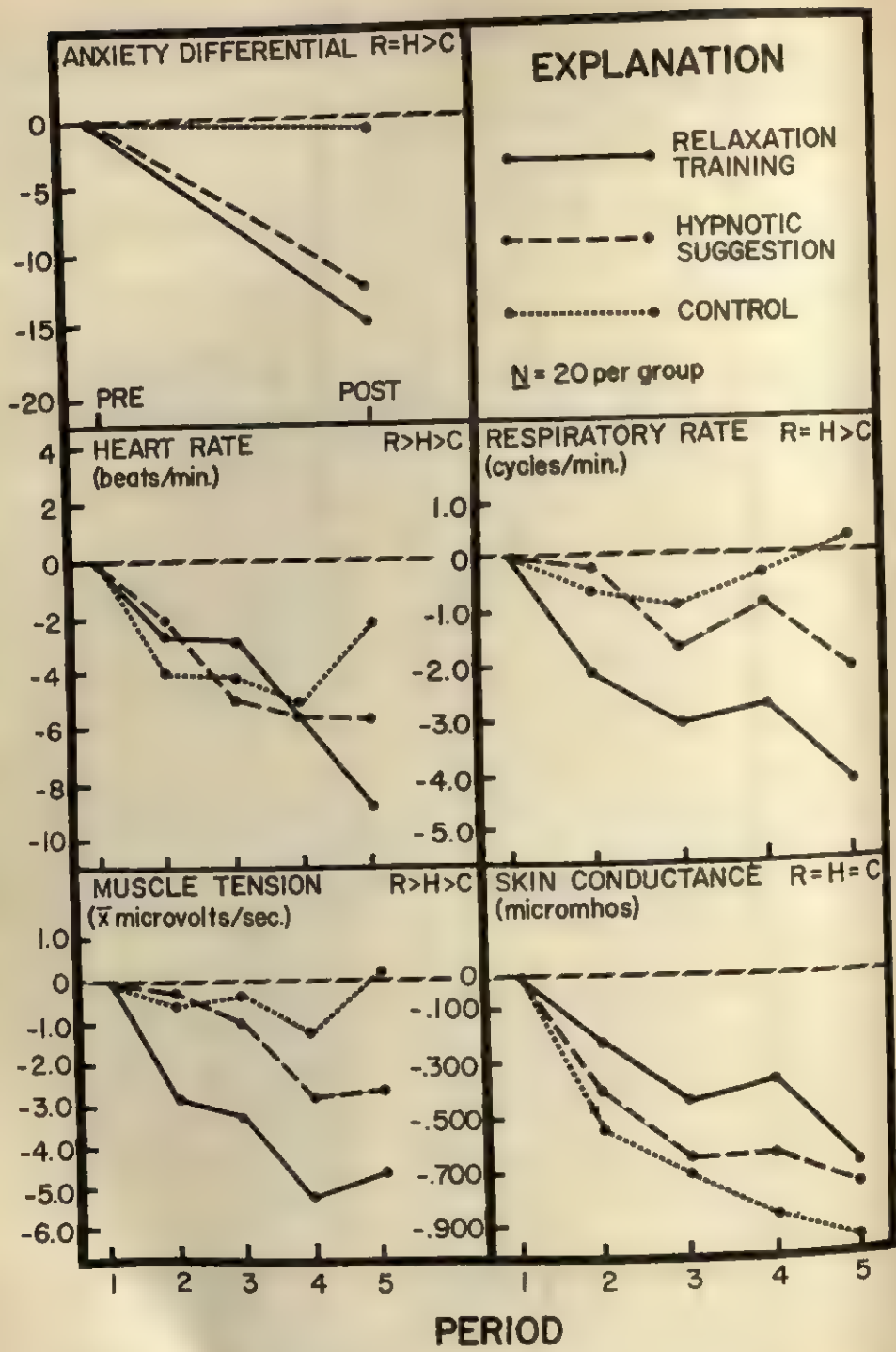


FIG. 2. Mean change from initial resting level on all measures during Session II.

fore, residual change scores, or "part scores," consisting of the difference between the obtained Period 5 score and that predicted by linear regression from the Period 1 score, were computed for each *S* on all measures before further analyses were undertaken. The overall evaluation of differential treatment effects was then carried out by analyses of variance on final period part scores for each measure in each session.

Highly significant treatment effects ($p < .01$, $df = 2/57$) were found for the Anxiety Differential in both sessions ($F = 11.083$ and 8.921) and for all physiological measures, except skin conductance (Session I: $F = 12.423$, 8.208 , 5.716 , for heart rate, respiratory rate, and muscle tension, respectively; Session II: $F = 9.786$, 6.190 , and 6.688 , respectively). Although the combined change for skin conductance was significant in both sessions, no significant differences were found between treatments on either occasion ($F = 1.696$ and < 1 , for Session I and II, $df = 2/57$). These analyses indicate that differential degrees of reduction in physiological arousal and distress occurred among groups.

A finer grained analysis was then undertaken on all measures except skin conductance to determine the specific effectiveness of

relaxation training and hypnotic suggestion, and whether relaxation training produced effects which differed from those attainable with hypnotic suggestion. The results of these multiple comparisons are summarized in Table 2. As seen by joint inspection of Figure 1 and Table 2, in the first experimental session, relaxation training produced significantly greater changes in the relaxed direction than controls on all measures tested. Additionally, relaxation training produced significantly greater decreases in heart rate and tonic muscle tension than hypnotic suggestion, while hypnotic suggestion resulted in significant decreases in respiratory rate and Anxiety Differential over controls. While the latter two measures followed the same trend as others, showing greater decreases for the relaxation group than for the hypnotic group, these differences did not achieve statistical significance.

In the second experimental session (see Table 2, Figure 2), relaxation training was, again, found to produce significantly greater decreases than controls on all measures tested. As in the first session, significantly greater reductions in heart rate and tonic muscle tension were found for relaxation training than for hypnotic suggestion. In the second session, however, the hypnosis group was also

TABLE 2

DUNCAN MULTIPLE-RANGE TESTS OF SIGNIFICANCE OF BETWEEN-GROUP DIFFERENCES ON FINAL PERIOD PART SCORES FOR EACH MEASURE

Measure	Comparison ($N = 20/\text{group}$)					
	Relaxation vs. control		Relaxation vs. hypnosis		Hypnosis vs. control	
	Obtained	SSR	Obtained	SSR	Obtained	SSR
Session I						
Anxiety Differential	14.311	> 5.674	2.812 (<i>ns</i>)	5.381	11.499	> 5.381
Heart rate	5.464	> 2.358	6.575	> 2.492	1.111 (<i>ns</i>)	2.358
Respiratory rate	3.834	> 1.700	1.293 (<i>ns</i>)	1.609	2.541	> 1.609
Muscle tension	3.462	> 2.031	3.267	> 1.922	.195 (<i>ns</i>)	1.922
Skin conductance*	—	—	—	—	—	—
Session II						
Anxiety Differential	10.715	> 3.620	1.039 (<i>ns</i>)	3.426	9.676	> 3.426
Heart rate	6.847	> 2.757	2.615	> 2.609	4.232	> 2.609
Respiratory rate	3.632	> 1.837	1.415 (<i>ns</i>)	1.738	2.217	> 1.738
Muscle tension	4.836	> 2.337	2.602	> 2.211	2.234	> 2.211
Skin conductance*	—	—	—	—	—	—

Note.—SSR = shortest significant range at $p < .05$ and joint protection level of 90%.

* Nonsignificant F ; therefore, multiple comparisons not computed.

TABLE 3

PERCENTAGE OF SUBJECTS SHOWING SIGNIFICANT REDUCTIONS IN THE RELAXED DIRECTION OVER COMBINED MEASURES

Treatment	No. measures ^a			
	0-1/4	2/4	3/4	4/4
Session I ^a				
Relaxation training		5	10	85
Hypnotic suggestion		20	50	30
Control	5	35	50	10
Session II ^b				
Relaxation training		10	20	70
Hypnotic suggestion		10	65	25
Control	35	35	25	5

Note.—Skin conductance not included as all Ss showed significant reductions.

^a $\chi^2 = 26.12$, $df = 4$, $p < .001$, $N = 20$ per treatment.

^b $\chi^2 = 35.90$, $df = 4$, $p < .001$, $N = 20$ per treatment.

found to produce significantly greater decreases than controls on all measures tested, suggesting a training effect. Overall comparative changes from the beginning of Session I to the end of Session II were identical with those found in Session II, with even greater between-groups differences in magnitude of response. Thus, it seems clear that both hypnotic suggestion and relaxation training did produce significant cognitive and physiological changes within the brief training sessions, and that relaxation training resulted in effects over and above those attributable to hypnotic suggestion.

As is typical of physiological data, significant positive but low intercorrelations (.20s and .30s) were found among part scores on at least one occasion for all measures, with the exception of skin conductance. All correlations between the latter part scores and other measures were negative, both within and between occasions, with about half reaching statistical significance ($|r| > .25$, $p < .05$). In contrast, the part-score stability coefficients for each measure from Session I to Session II were higher and all significant ($r = .76, .62, .69, .62$, and $.58$, respectively, for Anxiety Differential, heart rate, respiratory rate, muscle tension, and skin conductance, $p < .01$). These data suggest that the "relative response stereotypy" (Lacey & Lacey, 1958) frequently found in stress research, might also operate when physiological arousal is decreased as well as increased. However, excluding skin

conductance, highly significant coefficients of concordance (Siegel, 1956) were obtained for the relative consistency across measures of S response ($W = .37$ and $.30$ for Sessions I and II, $p < .001$) and relative treatment effect ($W = .81$ and 1.00 for Session I and II, $p < .01$).

While the magnitude of response on any given measure may be partially a function of each S's relative reactivity in the specific systems, the direction of change in each system should be the same such that differences in the degree of decreased arousal should also appear in the number of systems showing change in the relaxed direction, irrespective of magnitude. These data are summarized in Table 3, where the comparative figures on number of measures agree entirely with the earlier analyses on magnitude.

DISCUSSION

The combined findings show that both the brief relaxation training previously used with successful systematic desensitization therapy, and, the hypnotic suggestion procedure were effective in reducing subjective reports of tension and distress, even within a single session. By the second induction, hypnotic suggestion was found to produce significantly greater decreases in physiological arousal than self-relaxation controls, as measured by changes in heart rate, muscle tension, and respiratory rate; although, only the system under voluntary control (respiration) showed such effects during the first induction. In contrast, relaxation training was found to result in significantly greater decreases than controls on all of the above measures from the first training session, with effects maintained in the second. Further, the relaxation training procedure produced significantly greater reductions than hypnotic suggestion in systems not under direct voluntary control (heart rate, tonic muscle tension) in both sessions. Thus, with regard to the questions which led to the present study, it is clear that (a) both hypnotic suggestion and abbreviated progressive relaxation training do produce decreased physiological arousal and subjective distress within one to two sessions; (b) progressive relaxation training is more

effective than hypnotic suggestion in producing desired physiological changes, whether considered in terms of efficiency, intensity, or extent.

The findings with skin conductance, in which all Ss showed reductions over time in both sessions, without between-groups differences raise important questions regarding this measure. Higher conductance levels have recently been found under conditions of attention, novelty, or unfamiliarity than under actual stress (Flanagan, 1967; Roessler, Burah, & Childers, 1966), in some cases changing in the opposite direction to other physiological measures (MacNeilage, 1966), thus suggesting greater independence from sympathetic control. While autonomically mediated sweat gland activity has been assumed, historically, to form the primary basis for tonic changes, a number of factors have now been found to influence skin conductance (Darrow, 1967; Martin & Venables, 1966; Wilcott, 1967). Especially important is recent evidence that the permeability, or "water load," of the cornium and underlying layers of the epidermis between the surface electrode and deeper tissues influences the level of conductance, particularly under conditions of "low reactivity at high resistance," such as those of the present study (Darrow, 1964; Edelberg, 1966). Additional studies varying local blood supply by drugs or arterial occlusion indicate that, in the relative absence of sweat gland activity, increased vasodilation results in increased water load, and as a consequence, higher skin conductance (see Juniper, Blanton, & Dykman, 1967). Thus, it seems possible that skin conductance might gradually decrease along with sweat gland activity; however, as greater reductions in sympathetic activity occur, peripheral vasodilation could paradoxically effect the level of skin conductance. This hypothesis seems sufficiently tenable at present to recommend extreme caution in the interpretation of skin conductance in any investigation expected to reduce physiological activity below normal resting levels. Until additional clarification regarding skin conductance has been obtained, future investigations of the latter type might gain more relevant data through direct measurement of sweat production and blood volume.

Hypotheses regarding the influence of factors noted by Gill and Brenman and Barber and Calverley, described earlier, appear on examination of intrasession trends (Figures 1 and 2) combined with the significant differences in extent and magnitude of response. Such examination suggests that the limitations on sensory input and bodily activity plus motivational instructions, and eye closure—common factors in all three conditions—were not actively involved in reducing arousal level. These common factors may have been sufficient to avoid an increase in arousal, but apparently exerted little further effect after the adaptation period. Restricted attention, narrow and monotonous stimulus input, and suggested ease of response were additional factors common to both relaxation training and hypnotic conditions. The results obtained on the Anxiety Differential suggest that the latter factors in combination with either direct or indirect suggestions of relaxation, are sufficient to alter the cognitive experience of distress, while the results on physiological measures suggest that the differential content of stimulus input and attentional focus were of prime importance in the alteration of somatic responses. Thus, defining the situation as hypnosis, followed by direct suggestions of sleep, drowsiness, relaxation, etc., and arm immobility challenge, resulted in erratic physiological responses during the first session, with ultimate changes only in the voluntarily controlled response systems. During the second session, physiological response became less erratic, with ultimate change in systems not under direct voluntary control as well. On the other hand, defining the situation as passive learning, followed by systematic tension release of gross muscle groups, with indirect suggestions contingent upon Ss' own behavior, resulted in consistent response with ultimately greater changes in systems not under direct voluntary control in both sessions.

The findings for hypnotic suggestion add support to the growing body of evidence that hypnotic phenomena represent nothing discontinuous from other forms of behavioral influence. Hypnotic studies frequently report greater changes on self-report measures and voluntarily controlled response systems (Barber, 1965). These data are in agreement with

the interpretation of hypnotic effects within a learning framework in which suggestions may first lead to voluntary compliance, but, as restricted attention results in operator-emitted stimuli becoming more central and relevant, previously learned responses would be more likely elicited upon continued presentation of the verbal stimuli (Paul, 1963). In contrast to the hypnotic procedure, relaxation training appears to have induced changes as a consequence of abrupt release after tensing muscle systems, without relying solely on past learned responses to verbal or ideational stimuli. The indirect suggestions used in focusing attention on each muscle group during the release phase might, then, actually strengthen the association between these stimuli and physiological responses, such that verbal and ideational stimuli could become more adequate elicitors of relaxation responses, even in the absence of prior tension, as well as contributing to alter cognitive experience (see Paul, 1966; Paul & Shannon, 1966).

The results obtained in the present investigation reveal brief relaxation training to produce subjective and physiological changes in the direction opposite to anxiety, as assumed by the reciprocal inhibition principle from which systematic desensitization therapy was derived. A necessary requirement for interpreting the major mode of operation of systematic desensitization as response suppression or modified counterconditioning is thus met. Grossberg and Wilson (1967) have recently documented increased physiological response to the imaginal presentation of fearful stimuli, as compared to neutral stimuli appropriately counterbalanced across Ss for activity. Additionally, Davison (1968) found significantly greater improvements for Ss treated by systematic desensitization than for Ss treated by the same procedures without relaxation training. Several studies (Lang, 1969) have also reported fewer "signals" of anxiety during image presentation following relaxation training than in the absence of such training. Thus, while not excluding the importance of other factors, several converging lines of evidence support the contention that relaxation training produces effects which operate antagonistically to inhibit anxiety responses, as the underlying principle predicts.

REFERENCES

- BARBER, T. X. Experimental analyses of "hypnotic" behavior: A review of recent empirical findings. *Journal of Abnormal Psychology*, 1965, 70, 132-154.
- BARBER, T. X., & CALVERLEY, D. S. Empirical evidence for a theory of hypnotic behavior: Effects on suggestibility of five variables typically included in hypnotic induction procedures. *Journal of Consulting Psychology*, 1965, 29, 98-107.
- BARBER, T. X., & HAHN, K. W., JR. Hypnotic induction and "relaxation." *Archives of General Psychiatry*, 1963, 8, 295-300.
- BENJAMIN, L. S. Facts and artifacts in using analysis of covariance to "undo" the law of initial values. *Psychophysiology*, 1967, 4, 187-206.
- CAUTELA, J. Hypnosis and behavior therapy. *Behaviour Research and Therapy*, 1966, 4, 219-224.
- CORLEY, J. B. Hypnosis: A tranquilizer of choice. *American Journal of Clinical Hypnosis*, 1964, 6, 345-347.
- DARROW, C. W. The rationale for treating the change in galvanic skin response as a change in conductance. *Psychophysiology*, 1964, 1, 31-38.
- DARROW, C. W. Problems in the use of the galvanic skin response (GSR) as an index of cerebral function: Implications of the latent period. *Psychophysiology*, 1967, 3, 389-396.
- DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, 73, 91-99.
- EDELBERG, R. Response of cutaneous water barrier to ideational stimulation. *Journal of Comparative and Physiological Psychology*, 1966, 61, 28-33.
- FLANAGAN, J. Galvanic skin response: Emotion or attention? *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, 2, 7-8.
- GILL, M. M., & BRENNAN, M. *Hypnosis and related states*. New York: International Universities Press, 1959.
- GROSSBERG, J. M., & WILSON, H. Physiological changes accompanying imaginal fear situations. Paper presented at the meeting of the Western Psychological Association, San Francisco, May 1967.
- HILGARD, E. R. Individual differences in hypnotizability. In J. E. Gordon (Ed.), *Handbook of clinical and experimental hypnosis*. New York: Macmillan, 1967.
- HUSEK, T. R., & ALEXANDER, S. The effectiveness of the Anxiety Differential in examination situations. *Educational and Psychological Measurement*, 1963, 23, 309-318.
- JUNIFER, K., JR., BLANTON, D. E., & DYKMAN, R. A. Palmar skin resistance and sweat-gland counts in drug and non-drug states. *Psychophysiology*, 1967, 4, 231-243.
- KLINE, M. V. A visual imagery technique for the induction of hypnosis in certain refractory subjects. *Journal of Psychology*, 1953, 35, 227-228.
- LACEY, J. I., & LACEY, B. C. Verification and extension of the principle of autonomic response stereo-

- typy. *American Journal of Psychology*, 1958, 71, 50-73.
- LANG, P. J. The mechanics of desensitization and laboratory studies of human fear. In C. M. Franks (Ed.), *Assessment and status of the behavioral therapies and associated developments*. New York: McGraw-Hill, 1969.
- MACNEILAGE, P. F. Changes in electroencephalogram and other physiological measures during serial mental performance. *Psychophysiology*, 1966, 2, 344-353.
- MARTIN, I., & VENABLES, P. H. Mechanisms of palmar skin resistance and skin potential. *Psychological Bulletin*, 1966, 65, 347-357.
- PAUL, G. L. The production of blisters by hypnotic suggestion: Another look. *Psychosomatic Medicine*, 1963, 25, 233-244.
- PAUL, G. L. Supplementary notes on the clinical use of hypnosis. Unpublished course supplement, University of Illinois, 1965.
- PAUL, G. L. *Insight vs. desensitization in psychotherapy: An experiment in anxiety reduction*. Stanford: Stanford University Press, 1966.
- PAUL, G. L. Insight versus desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology*, 1967, 31, 333-348.
- PAUL, G. L. A two-year follow-up of systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, 1968, 73, 119-130.
- PAUL, G. L. Outcome of systematic desensitization. I. Background, procedures, and uncontrolled reports of individual treatment. In C. M. Franks (Ed.), *Assessment and status of the behavioral therapies and associated developments*. New York: McGraw-Hill, 1969. (a)
- PAUL, G. L. Outcome of systematic desensitization. II. Controlled investigations of individual treatment, technique variations, and current status. In C. M. Franks (Ed.), *Assessment and status of the behavioral therapies and associated developments*. New York: McGraw-Hill, 1969. (b)
- PAUL, G. L., & SHANNON, D. T. Treatment of anxiety through systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, 1966, 71, 124-135.
- ROESSLER, R., BURAH, N. R., & CHILDERS, H. E. Personality and arousal correlates of specific galvanic skin responses. *Psychophysiology*, 1966, 3, 115-130.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- SPIELBERGER, C. D. Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press, 1966.
- TUCKER, L. R., DAMARIN, F., & MESSICK, S. J. A base-free measure of change. *Psychometrika*, 1966, 31, 457-473.
- WEITZENHOFFER, A. M. *General techniques of hypnosis*. New York: Grune & Stratton, 1957.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Forms A and B*. Palo Alto: Consulting Psychologists Press, 1959.
- WILCOTT, R. C. Arousal sweating and electrodermal phenomena. *Psychological Bulletin*, 1967, 67, 58-72.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.

(Received March 25, 1968)

EFFECTS OF VARIATIONS IN THE PREPARATORY INTERVAL ON THE REACTION TIMES OF RETARDATES AND NORMALS¹

ALFRED A. BAUMEISTER¹ AND STEPHEN J. WILCOX²

University of Alabama

Two studies were conducted to evaluate the effects, on simple reaction times, of warning signal duration (WSD) and preparatory interval (PI) with normal and mentally retarded Ss. In the first study PIs of 2, 4, 8, and 12 sec. were presented in an unpredictable sequence. For half of the Ss the warning signal remained on throughout the entire PI (filled condition) and for the remainder of the Ss, the warning signal was present for only 1.5 sec. of the PI. Significant effects of intelligence, PI, and WSD were found. The interaction Intelligence \times PI was also significant. The unfilled condition produced faster RTs for both normals and retardates. The interaction resulted from the particularly poor performance of the retardates at the short PIs. The second study employed a 1.5-sec. WSD with PIs of 2, 4, 8, 16, and 32 sec. An interaction again was found involving intelligence and PI. The results were discussed in terms of temporal uncertainty.

The procedures employed in the reaction-time (RT) task typically involve presentation of a discrete warning signal followed by a blank interval. The entire period set off from the beginning of the warning signal to the appearance of the reaction signal is known as the preparatory interval (PI). An implicit response sequence, sometimes referred to as "preparatory set," is assumed to occur during PI as S readies himself to respond quickly to the reaction stimulus.

Usually such "sets" are indirectly inferred from the effects that manipulations of the PI produce upon reaction time. If the intervals within a session are unpredictably variable in duration and S is therefore uncertain about the onset of the reaction signal, he must maintain his readiness throughout the entire interval in order to be consistently fast in his responding. The ability to maintain this set has been the focus of a number of researchers working with abnormal persons. One implication of this concern is that certain types of disturbances are reflected in an impaired capacity to hold or direct attention.

Recently several studies have been con-

ducted in which the RTs of mental retardates and normals have been compared in relation to certain manipulations of the PI. Hermelin (1964) has reported that subnormal Ss are differentially disadvantaged by long PIs. An interaction of this nature is predicted by Ellis (1963) who has postulated that retardates suffer short-term memory (STM) defects and, consequently, are particularly handicapped in situations that require them to respond on the basis of absent cues. A study by Terrell and Ellis (1964), conducted as a direct test of this STM deficit hypothesis, compared normal and retarded Ss under conditions where PI was either partially or completely filled by the warning signal. Their data indicate an interaction in which the retardates, but not the normals were facilitated by the "filled" intervals. Supposedly, dependency on STM was less critical when the warning signal was constantly present. Unfortunately, a number of more recent studies, reviewed by Baumeister and Kellas (1968), not only fail to confirm the findings of Hermelin (1964) and Terrell and Ellis (1964) but actually report results that lead to opposing conclusions (e.g., Hawkins & Baumeister, 1965; Kellas & Baumeister, 1968).

Procedural differences among the various studies are of sufficient magnitude to make

¹Requests for reprints should be sent to Alfred A. Baumeister, Department of Psychology, University of Alabama, University, Alabama 35846.

²The authors wish to express their gratitude to the patients and staff at Partlow State Hospital and School, Tuscaloosa, Alabama, for their cooperation.

direct interexperimental comparisons virtually impossible. Any one of a number of procedural factors conceivably could account for the discrepant findings. The two experiments reported here represent an effort to evaluate further the nature of possible interactions involving characteristics of PI and intelligence. The first study is a fairly close replication of the procedures employed by Terrell and Ellis, the second investigation was concerned with the contextual effects of irregular PIs differing markedly in duration.

EXPERIMENT I

Method

Subjects. An effort was made to obtain Ss similar to those tested by Terrell and Ellis (1964). Thirty Ss between the ages of 17 and 28 were selected from the population of a state institution for the mentally retarded. The mean age of the group was 22.2 yr., $SD = 3$. Their IQs, taken from institution records, had a mean of 55.5, $SD = 7.2$. Ss exhibiting gross sensory or motor defects were not included in the sample. A comparison group of 30 normals was selected from an introductory psychology class at the University of Alabama. Their mean age was 20.4, $SD = 5.1$.

Apparatus. The apparatus consisted of a red warning light, 1 in. in diameter, a response telegraph key, a door buzzer, a Hunter Klockounter, and a series of decade timers for control of stimulus presentation. The light and response key were mounted on a small table. The retardates were tested in a small room on the institution grounds, the normals in a similar room at the University. Both rooms were dark except for illumination provided by a single 7-w. lamp. Recording and controlling equipment was located in adjacent rooms.

Procedure. Normals and retardates were randomly divided into two subgroups of 15 Ss each. Tests of significance revealed no chronological age (CA) bias in the assignment of Ss to subgroups. No significant difference was found between average IQs of the retarded subgroups.

One subgroup received a *filled* condition in which the warning light remained on throughout the PI until the response was made to the reaction signal (door buzzer). The other group received an *unfilled* condition in which the warning light was presented only for the initial 1.5 sec. of the PI. All Ss were given 10 trial blocks consisting of 4 trials each. In every trial block each PI of 2, 4, 8, and 12 sec. was presented in random order. Each S was tested with a different random sequence of PIs. The inter-trial interval was 8 sec.

The S was instructed to depress the response key at the onset of the warning light and to hold it down until the buzzer sounded. He was told to lift his finger as quickly as possible upon hearing the buzzer.

The swiftness of the response was emphasized in the instructions. Three practice trials were given. If necessary, the instructions were repeated or clarified during these trials. Before entering the test room, the retardates were shown an assortment of prizes and were told that if they "did well" they could choose one to keep after the test.

Results

Medians were obtained for the 10 RTs for each S under the various warning conditions. Means of these median scores are plotted in Figure 1. Standard deviations in relation to these points ranged .03-.08 sec.^a The results of analysis of variance applied to the median data are summarized in Table 1.

Main effects of PI, intelligence, and warning signal duration (WSD) were all significant. With regard to the WSD effect, it should be emphasized that it was the *unfilled* condition that produced the fastest reactions for both normals and retardates ($p < .01$).

The interaction Intelligence \times PI was also significant. Simple analyses of this effect revealed that the retardates, relative to the normals, were differentially affected ($p < .05$) at the *shortest* (2-sec.) PI. An examination of the sequential pattern of the warning periods indicated that the retardates were particularly slow on trials in which the 2-sec. PI had been preceded by the longest (12-sec.) interval.

EXPERIMENT II

One point on which the present findings are in agreement with those of Terrell and Ellis concerns the interactive effects of PI and intelligence on RT. In both studies the retardates were relatively disadvantaged under the shortest PIs. This result directly contradicts expectations based on the stimulus-trace hypothesis and the findings of Hermelin (1964).

The second study in this series was undertaken to assess further the effects of the PI on the RTs of normal and retarded Ss. In

^a Differences in variance between intelligence groups, while not extreme, were nevertheless significant ($F_{\max} = 9.84$). Consequently, the critical values were obtained by referring to tabled values for one-half the alpha level actually reported. This is the procedure recommended by Lindquist (1953) to avoid understating the probability of a Type I error.

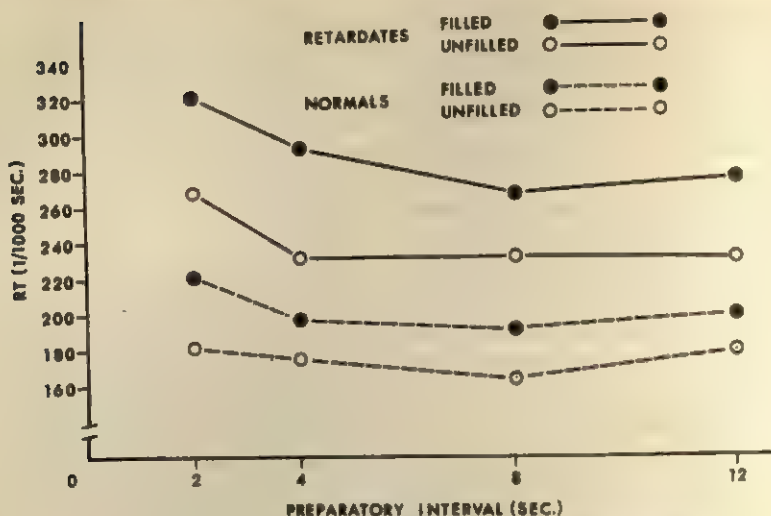


FIG. 1. RT as a function of intelligence group, preparatory interval, and warning condition (Study 1).

this case, some of the intervals were extended to 32 sec. in order to insure a fair test of the predicted interaction Intelligence \times Delay Interval. The retardates were subdivided into "fast" and "slow" groups on the basis of previous RT scores. This procedure was followed to permit an evaluation of the effects of PI in relation to general RT ability within the retarded population.

Method

Subjects. The normals were 15 volunteer college students from an introductory psychology class. Their mean age was 20.6 yr. The retardates were 30 institutionalized males who had participated previously in at least two RT experiments. On the basis of their earlier performance, the retarded Ss were assigned equally to *fast* and *slow* groups. The mean ages were 22.9 yr. and 23.2 yr.; the mean IQs were 60.31 and 51.3, respectively. The difference

in mean IQ was found to be statistically significant ($p < .05$).

Apparatus. The apparatus and test conditions were the same as in Experiment I.

Procedures. Procedures were essentially identical to those employed in the first study under the unfilled condition. A warning signal of 1.5 sec. was followed by a variable interval of 2, 4, 8, 16, or 32 sec. These intervals were presented systematically over 50 trials so that each PI preceded every other PI an equal number of times. All Ss within each of the three subgroups received different sequences of PIs.

Results

The 10 observations for each S under each of the preparatory conditions were averaged. Curves for the three groups are displayed in Figure 2. Standard deviations ranged .03–.10 sec.⁴

Analysis of variance based on these data revealed the interaction ($p < .01$) Groups \times PI. An inspection of Figure 2 discloses that this interaction arose from the relatively poor performance of the slow group of retardates at the 2-sec. PI. A more detailed examination of the contextual effects of the various PIs indicated that although all three groups showed a tendency to respond more slowly on 2-sec. trials that had been immediately preceded by trials involving a long PI, the slow retardates

TABLE 1
ANALYSIS OF VARIANCE ON MEDIAN REACTION TIMES
FOR STUDY 1

Source	df	MS	F
Intelligence (A)	1	339.077	37.7***
Filled-unfilled (B)	1	84.113	9.36**
A \times B	1	6.731	.74
Error (A, B, A \times B)	56	8.986	
Preparatory interval (C)	3	13.765	24.7
A \times C	3	2.414	4.31*
B \times C	3	.812	1.5
A \times B \times C	3	.765	1.4
Error (C, A \times C, B \times C, A \times B \times C)	168	.557	

* $p < .025$.

** $p < .01$.

*** $p < .001$.

⁴As in Experiment I, the intelligent groups were not equally variable in performance. Again, an adjustment was made in obtaining the critical values.

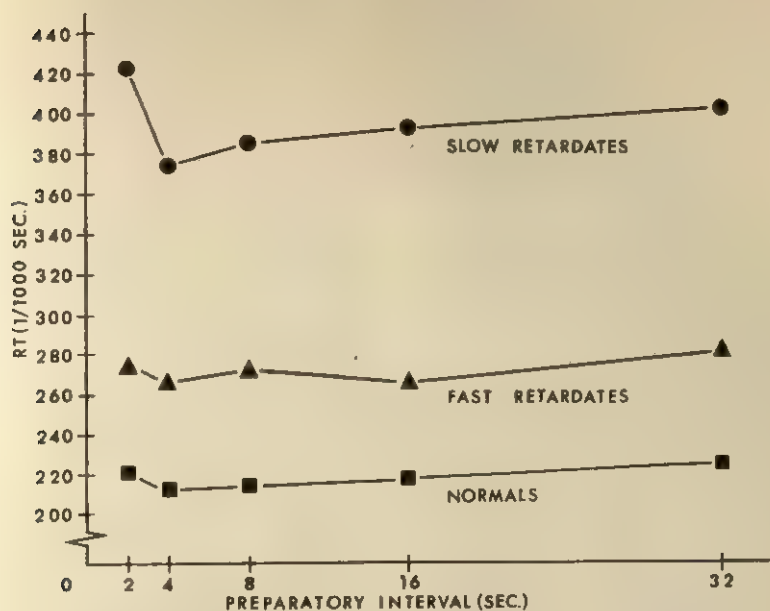


FIG. 2. RT as a function of group and preparatory interval (Study 2).

were particularly disadvantaged under this condition.

DISCUSSION

To the extent that the variables manipulated in this series of studies reflect the operation of STM processes, one may conclude that no evidence has been found to support speculation that mental retardates suffer a STM deficit. Specifically, Ellis (1963) has hypothesized that retardates would benefit, relative to normals, in an RT situation where the warning signal was continuously available during the PI. Terrell and Ellis (1964) did report an interaction of this nature, a finding which the results of the present investigation substantially contradict.

Because the procedures employed here were highly similar to those used by Terrell and Ellis, it is difficult to identify the sources of the discrepancy. The minor procedural differences which are apparent simply do not seem sufficient to account for the completely opposite findings concerning the effects of the WSD.

It is obvious that the two warning signal conditions, filled and unfilled, did not provide the same information to the S concerning the onset of the reaction signal. Uncer-

tainty may have been greater when the PI was completely filled by the warning light. Under this condition, S was required to ready himself from the beginning of the interval. On the other hand, the unfilled warning condition afforded an additional cue with the cessation of the warning signal. In this case, S did not have to prepare for the reaction stimulus until the termination of the warning signal. The period during which S was required to maintain his state of readiness was thus shorter under the unfilled condition. If this interpretation is correct, one may expect that the difference between the two warning conditions would be most pronounced at the shortest interval, where the warning signal itself constituted a substantial portion of the preparatory period. An analysis based on mean RTs, in fact, does indicate ($p < .025$) that the difference in RTs between the two warning conditions was greatest at the short PIs.

Although an interaction was found involving intelligence groups and length of PI, the effect, again, was opposite to that predicted by the STM position of Ellis (1963). In this case, the short intervals produced the greatest differences between normals and retardates. It should be noted, also, that these data

contradict Hermelin (1964) who found that subnormals were more disadvantaged than normals by progressive lengthening of the PI. However, the results of the present studies are fairly consistent with the majority of investigations concerning the effects of irregular preparatory periods on RT.

The finding that the slow retarded group (Study 2), in particular, was adversely affected at the short intervals suggests that these Ss had difficulty in learning the statistical characteristics of PI. A trial-by-trial examination of the data indicated these Ss were especially slow on 2-sec. trials that were immediately preceded by the long intervals (16 or 32 sec.). It is as if S is led to "expect" a long interval and is, therefore, caught unprepared when the interval is very brief. While this context effect has been reported previously for both normals (e.g., Klemmer, 1956) and retardates (e.g., Baumeister & Hawkins, 1966), it does appear to be more pronounced among less adaptive individuals.

In view of the findings concerning the effects of various temporal manipulations of the warning interval in the typical simple RT task, it seems more meaningful to invoke an explanatory concept such as "expectancy" or set rather than STM. For one thing, the operations employed to define STM in this context are unclear (e.g., exactly what is S supposed to remember and over what inter-

val?) and, for another, the effects of these manipulations are inconsistent with predictions generated within a STM position.

REFERENCES

- BAUMEISTER, A. A., & HAWKINS, W. F. *American Journal of Mental Deficiency*, 1966, 70, 689-694.
- BAUMEISTER, A. A., & KELLAS, G. Reaction time and mental retardation. In N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol. 3. New York: Academic Press, 1968.
- ELLIS, N. R. The stimulus trace and behavioral inadequacy. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.
- HAWKINS, W. F., & BAUMEISTER, A. A. Effect of duration of warning signal on reaction times of mental defectives. *Perceptual and Motor Skills*, 1965, 21, 179-182.
- HERMELIN, B. Effects of variation in the warning signal on reaction times of severe subnormals. *Quarterly Journal of Experimental Psychology*, 1964, 16, 241-249.
- KELLAS, G., & BAUMEISTER, A. A. The effects of warning signal duration on reaction times of mental defectives. *American Journal of Mental Deficiency*, 1968, 72, 668-673.
- KLEMMER, E. T. Time uncertainty in simple reaction time. *Journal of Experimental Psychology*, 1956, 51, 179-184.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- TERRELL, C. G., & ELLIS, N. R. Reaction time in normal and defective subjects following varied warning conditions. *Journal of Abnormal and Social Psychology*, 1964, 69, 449-452.

(Received March 28, 1968)

SUBJECTIVE REPORT AND CREDIBILITY: AN INQUIRY INVOLVING HYPNOTIC HALLUCINATIONS¹

KENNETH S. BOWERS² AND J. BARNARD GILMORE³

University of Waterloo

This paper argued that credibility of subjective reports is in part situationally determined. An experimental demonstration was considered showing that under conditions where report credibility was a reasonable assumption, hypnotic Ss reported previously suggested hallucinations, whereas simulator controls did not. The experimental evidence also suggests that hypnotic Ss are honest and their reports credible across a variety of situations. It was concluded that subjective reports of hypnotic Ss may well correspond to actual perceptual experience, and that motivated compliance does not satisfactorily explain hypnotic phenomena.

Science has frequently been able to minimize its reliance on data which are said to be nonobjective, but no science has entirely circumvented subjective experience. In psychology, it often seems that the richest sources for understanding human behavior are the most purely subjective ones. Hypnotic phenomena are a striking case in point. They appear to offer access to processes involved in perception, cognition, attitudes, fantasy, and defenses. Yet hypnotic phenomena, being highly subjective in nature, are often dismissed or ignored by psychologists. The emphatically subjective quality of the hypnotic experience makes it suspect, and even incredible, to those of a behavioristic persuasion.

Simple motivational and/or behavioristic explanations of hypnotic behavior have been attempted (Barber & Calverley, 1962; Edmonston, 1967; Hull, 1933). However, what hypnotic Ss say about their experiences often does not seem to jibe with such explanations. It is possible, of course, that what hypnotic Ss say about their experiences is false. Consequently, it is not unreasonable to ask whether and when persons tell the truth

about their experiences. It is certainly the case that the validity of any report presuming to reflect subjective experience cannot be directly evaluated. However, in everyday life we do tend to value certain kinds of evidence—including verbal reports—as pertinent to drawing conclusions about another person's subjective experience. For example, we ordinarily accept as credible the report of a nightmare by a child who has awakened crying out in terror for his mother, even though the dream itself was inaccessible to direct observation. There are other occasions, however, when we do not take a person's remarks at face value. Flattery is often suspect, as are the nostrums of politicians seeking office. Thus, there are factors implicit in any given context that help determine what is credible and what is not.

It is not the intention of this paper to investigate the parameters of credibility *per se*. Bem (1965) and Jones and Davis (1965) have made constructive attempts to this end. Rather, the authors would like to state an intuitively plausible hypothesis about report credibility, and to use this assumption systematically in studying a highly subjective phenomenon, namely, hypnotic hallucinations.

The working hypothesis is that a person's experiential report is credible in inverse proportion to the extent that this report is demanded and/or contingently reinforced by the environment. Given such an assumption it is the authors' hope that the methods and reasoning employed in the present experiment

¹ This study was supported by Research Grant No. 100 from the Ontario Mental Health Foundation and by Research Grant No. MA-2328 from the Medical Research Council.

² Requests for reprints should be sent to Kenneth S. Bowers, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

³ The authors thank Sandra van der Meulen for collecting the data of this experiment, and for contributing her insights into some of the procedural problems the authors had to overcome.

will have general significance for solving some current problems in experimental hypnosis, and indeed for all psychological investigations in which Ss' verbal reports constitute prime or ancillary data.

Consider, then, the case of hypnotically induced hallucinations. When does an S's report of an hypnotically induced hallucination reflect his actual experience, and when does it serve only to please the operator by complying with his expectancies and the general situational demands? Put another way, under what conditions are subjective reports apt to reflect or correspond to an S's hallucinatory experience, and under what conditions are they likely to be a function of S's perception of what he is *supposed* to say in the situation, irrespective of his experience?

An answer to such questions, in terms of the assumption concerning report credibility, is fairly simple. Reports of hypnotic hallucinations, given in a context free of demands to report them, are more credible than are similar reports given in a context where hallucinatory reports are implicitly demanded or reinforced.

However, translating this principle into a concrete demonstration presents an immediate practical problem: Since suggestions to hallucinate constitute implicit demands to report hallucinations, how can one divest a report context of implicit demands to report the hallucinations suggested? It was to get at this issue that Bowers (1967) partially replicated an earlier study by Barber and Calverley (1964). The latter authors had shown that Ss receiving hypnotic suggestions to hallucinate and Ss task motivated to hallucinate both rated their hallucinations as significantly more real than control Ss. Bowers (1967), however, pointed out that in both the hypnotic and task-motivated groups, the experiential reports of hallucinations were given in a context implicitly demanding reports of hallucinations. That is, Ss were first given suggestions to hallucinate, and then immediately asked whether they in fact hallucinated. Implicit in the initial suggestions was a demand to report hallucinations. Thus, although Ss' reports may have reflected an actual hallucinatory experience, it is equally and perhaps even more likely that the reports were a

function of situational demands to report hallucinations. The existence of the latter possibility seriously vitiates report credibility.

To obviate this ambiguity, Bowers (1967) ran two groups of task-motivated Ss. The first group rated the reality of their hallucinatory experiences under the same set of demands that characterized the hallucinatory episode itself. The Ss in the second task-motivated group, however, before rating their experience, were confronted with (a) a demand for honesty in all subsequent reports, and (b) an interviewer who had previously been unconnected with the experiment. As in the Barber and Calverley (1964) study, Ss in the first group significantly increased their post-test ratings of the hallucinations' reality over a pretest level. The Ss in the second group showed no rating increase whatever. It was concluded from these findings that the context in which reports of hallucinations occur is very important. When the report context is purged of demands for high ratings (i.e., hallucinations rated as very real), Ss' ratings remained at relatively low pretest levels.

Although the methods of the present experiment are quite different from those employed in Bowers' (1967) study, the strategy is analogous. The question is the following: In a context free of situational demands to report hallucinations, what will hypnotic Ss report? If they report the hallucinations previously suggested and do so in a context which does not implicitly demand such reports, it can be argued that the reports are referable to their actual experience.

The demonstration in this experiment that a context is free of demands to report hallucinations hinges to some extent on the behavior of quasicontrol Ss selected for their unsusceptibility. It is assumed that unsusceptible Ss are incapable of experiencing hypnotically induced hallucinations. Consequently, reports of hallucinations by these Ss would suggest the presence of situational demands to report them. Conversely, the absence of hallucinatory reports by these Ss would imply the absence of demands to report them. Therefore, it is in a context in which these quasicontrol Ss do *not* report hallucinations that hallucinatory reports by hypnotic Ss become credible.

METHOD

Subjects

A total of 23 highly susceptible *Ss* and 14 unsusceptible *Ss* were selected for this experiment. The susceptible *Ss* had all been in previous hypnotic experiments conducted at the University of Waterloo, and were initially selected as being some of the highest scoring *Ss* on the Harvard Group Scale of Hypnotic Susceptibility (HGSS, Shor & Orne, 1962) and the individually administered Form C of the Stanford Hypnotic Susceptibility Scale (SHSS, Weitzenhoffer & Hilgard, 1962).⁴ Altogether, the susceptible *Ss* had been seen for at least six individual training and testing sessions. Of the highly susceptible *Ss*, 16 were assigned to the hypnotic condition, and 7 to a control condition in which they were asked to simulate hypnosis. The unsusceptible *Ss* were selected on the basis of a score of 0-3 on the HGSS and were also assigned to a simulating condition (0-3 simulators). The 0-3 simulators were used to minimize the possibility of their becoming inadvertently hypnotized during the course of the experiment (Hilgard & Tart, 1966; Tart & Hilgard, 1966). The susceptible simulators were employed to control for *S* effects, and for the effect of *S* familiarity with hypnotic techniques.

*Procedure*⁵

The *Ss* in the experiment were seen by a hypnotist who was unfamiliar with them and who therefore did not know in advance whether a particular *S* was susceptible or unsusceptible. Upon entering the laboratory, *S* was told to sit in a large, comfortable chair. The *E* then told *S* that earphones would be placed on his head, and taped instructions would be played. The *S* was told that *E* did not know which of several instructions would be played, as the particular tape he was about to hear was randomly selected by a third person. This in fact was true. Furthermore, *S* was told not to reveal to *E* by word or sign, the nature of the taped instructions. After placing the earphones on *S* and turning on the tape recorder, *E* left the room so as not to be present during the playback. In this way, any facial or other cues made by *S* during the instructions were unavailable to *E*, thereby tending to preserve *E*'s naïveté concerning *S*'s condition.

One kind of tape contained instructions to simulate hypnosis, and to fool the hypnotist into believing *S* was in fact hypnotized. The exact wording of the instructions was similar to those reported earlier (Bowers, 1966). The second kind of instructions merely told *S* that he would subsequently be hypnotized. Much of this recording consisted of total

⁴ The scores of the susceptible *Ss* on the HGSS ranged 8-12 and on the C scale 7-12.

⁵ The present experiment is actually the second of two experiments which were run one after the other, always in the same order. The results of the first experiment were completely negative and are not reported here.

silence, so as to make the tape equal in length to the simulator instructions. Both tapes contained explicit demands for honest reports in a post-experimental interview (PEI), after a signal from *E* ("Now the experiment is over") was given. The last instruction was to remove the earphones.

After *S* had removed the earphones, *E* returned and delivered hypnotic suggestions adapted closely from those used on the SHSS Form C. When *S* appeared to be deeply hypnotized, instructions and suggestions for the first experiment were delivered (see Footnote 5). After completing the first experiment, *E* introduced *S* to the second experiment. The materials for this experiment consisted of a series of 5 × 7 in. display cards, on all but two of which a color chip, approximately 1 in. on a side, appeared. Each chip consisted of colored, transparent cellophane mounted on top of two intersecting black lines at right angles to each other. The lines extended well beyond the boundaries of the chip, and were easily visible underneath it. The chips differed independently along three dimensions: color (red or green), shape (square or triangular), and position (the chips appeared in either the upper or lower half of the display card). Each *S* in the experiment was first exposed to the same two sample display cards, clearly labeled A and B, and presented one after the other. The hypnotized (or simulating) *S* was instructed to open his eyes and look at Display A. The *E* then described Display A for *S*, noting that a red square chip was in the upper section of the display card. After eliciting agreement from *S* about the color chip on Display A, *E* replaced Display A with Display B and told *S* to describe the color chip thereon. The *Ss* invariably described the chip accurately (green, triangular, lower section).

Next, *S* was presented with a pair of conjoined display cards labeled Display I and Display II, respectively. Each *S* was confronted with one such pair randomly selected from a pool of eight card pairs constructed for the experiment. The color chips on these eight pairs of display cards differed systematically and independently from each other. For example, Display I of one pair of cards might consist of a green triangle in the lower section and Display II might be a red triangle in the upper section. This pair of display cards, taken as a unit, differed from all other seven paired displays. The display cards were presented in a way that made it impossible for *E* to know which one of the eight different card pairs he was presenting to a given *S*. Thus, it was impossible for *E* to communicate to *S*, however unwittingly, any biases arising out of his knowledge of what color chips actually appeared on the displays.

Before actually opening his eyes to view Displays I and II, *S* was given further suggestions to deepen hypnosis. Then he was told:

First I'm going to show you Displays I and II. Each of these displays resembles Displays A and B which you have already seen, except for one thing. On neither Display I nor II will there be any

color chips. You will be able to see Displays I and II quite clearly, but on neither of them will you see *any* color chip whatsoever; Display Cards I and II will be absolutely empty. There will be no color chips in either of the two displays. Do you understand? O.K. Open your eyes slowly and look at the two blank display cards. What do you see on Display I? What do you see on Display II? O.K. Close your eyes for a few moments while I change the two displays. [S's answers were recorded.]

Several things should be noted at this point before going on. First, and obviously, the authors are here suggesting negative hallucinations. Second, *reports* of negative hallucinations are being elicited in a context which implicitly demands of *S* a report of "I don't see anything." If *S* wished only to comply verbally with the situational demands, he knows everything he needs to know about what to say. Moreover, if the suggestion for negatively hallucinating the color chips were successful in eliciting the hallucination, *S*'s verbal report to this effect, taken in this context, would not differentiate him from the merely compliant *S*.

The next step in the experiment was to present *S* with two conjoined display cards labeled Displays III and IV which had *no color chips* on them whatever. In every respect other than the differences in the labels (III and IV instead of I and II) and the absence of color chips, the cards were identical with those previously viewed. The *S* was then told:

Now I am going to show you the last pair of display cards, Displays III and IV. When I ask you to open your eyes in a few moments, you will see a color chip on both Display III and Display IV. You will see the two chips quite clearly just as you saw the color chips a while ago on Displays A and B. The chips will be either red or green, square or triangular, and in either the upper or lower section of each display. When you open your eyes, I want you to tell me exactly where you see the color chips, what colors they are, and what shape. Do you understand?

The *S* was then told to open his eyes and was asked for the (hallucinated) color, shape, and position of the chips on each of the two displays. After yielding his response, *S* was commended and instructed to close his eyes. All of *S*'s responses were recorded. Again notice that the demands inherent in the suggestion for positive hallucinations clearly imply the preferred kind of verbal report.

Next, *E* said the following to those *Ss* who had earlier reported negative hallucinations to Displays I and II.

Now, do you remember just a moment ago when you opened your eyes and saw nothing on Displays I and II? In fact, there was a color chip on each display then, but because of my suggestion you thought you did not see them. Now I want you to recall what in fact *was* there that first

time. This sounds hard, but you really did see the color chips and now that I am removing my earlier suggestion, you will find that you *can* remember seeing the chips that were there. You will find that my reminding you that something really was on each display helps you remember now seeing what was really there. Now you can remember what was really on Displays I and II.

These *Ss* were then asked to recall the chips the very presence of which were not acknowledged earlier. Those *Ss* who initially did identify the color chips on Displays I and II were simply asked to recall them again.

By suggesting recall for previously unacknowledged color chips, the authors attempted to alter the demands of the situation sufficiently so as to produce memories of what was previously "not seen." Orne (1962) has noted that *Ss* who have negatively hallucinated a chair "demonstrate that at some level they are aware of the chair" by tending to avoid it while walking around the room. One might expect therefore, that hypnotic *Ss* might be able, on command, to retrieve previously registered but heretofore unavailable information.

The *S* was then given posthypnotic suggestions to remember everything that had transpired during the experiment. Finally, *S* was awakened and immediately told "Now the experiment is over." It will be remembered that this phrase was introduced during the taped instructions as the signal for simulating *Ss* to stop faking, and for *Ss* in both conditions to answer all subsequent PEI questions as honestly as possible. All *Ss* were asked two initial questions: "[Question 1] Were you asked on the tape recorder to fake hypnosis, and to fool the hypnotist into believing that you were hypnotized (yes or no)? [Question 2] Whether or not you were asked to fake hypnosis, did you become hypnotized (yes or no)?"

Then followed the structured PEI, the aim of which was to elicit *S*'s experience of the test situation. The most important and revealing of these questions follow.

[Question 3] Would you say that the color chips you now remember seeing on Display Cards I and II were (a) less real, (b) equally as real, (c) more real than the color chips which appeared on the original Display Cards A and B?

[Question 5] For some people in this experiment, color chips appeared on Display Cards III and IV. For other people, there were no color chips on Displays III and IV. Were you in the condition in which (a) color chips were present on Displays III and IV, (b) color chips were not present on Displays III and IV?

The *Ss* who correctly stated that color chips did not appear on Displays III and IV were asked no further questions. The *Ss* who affirmed the presence of color chips on Displays III and IV were asked several more questions:

[Question 6] Would you say that the color chips which appeared on Display Cards III and IV were

(a) less real, (b) equally as real, or (c) more real than the color chips you now remember appearing on Display Cards I and II?

[Question 7] Would you say that the color chips which appeared on Display Cards III and IV were (a) less real, (b) equally as real, or (c) more real than the color chips which appeared on Cards A and B?

[Question 10] Which would now seem most incredible to you: that there really were no color chips on Displays Card I and II after all, or that there were no color chips on III and IV.

The significant aspect of the PEI reports of Ss is that they were given in a context which was hopefully free of demands to report hallucinatory experiences. In this respect, the PEI differs importantly from the report context of the experiment per se. Precisely because giving "correct" responses is presumably not an issue in the PEI, the reports made in this context are potentially more credible insofar as they are more apt to be referable to S's subjective experience.

The success of the manipulations to evoke honest reports can be assessed in part by the answers of the unsusceptible, 0-3 simulating Ss during the PEI. Because 0-3 simulators are essentially unhypnotizable, any reports of hallucinatory experiences they give during the PEI would be suspect as reflecting their response to residual demands implicit in the report context, rather than being an accurate experiential account. If it is at all common for 0-3 simulators to report hallucinations in the PEI, then the significance of similar reports by the hypnotic Ss would not be credible.

RESULTS

Subject Reports during Experiment

The importance of the design of this study is that it allows direct comparison of verbal reports of Ss in two conditions (hypnotic and simulating) in two different report contexts.^a According to the argument, mere verbal compliance is most apt to characterize Ss' reports during the experiment per se, where demands for hallucinations are implicit in the hallucinatory suggestions. Table 1 shows the breakdown of Ss' reports of negative hallucinations to Displays I and II during the experiment. Observation of this table indicates that one-half of the 0-3 simulators complied with the

^a Actually, there were four conditions, if the susceptible simulator Ss are included. Since the number of these Ss was relatively small, and because, with the exception of one S, the data of this group were very similar to that of 0-3 simulators, relatively little space is devoted to presenting the results of the susceptible Ss.

TABLE 1

REPORTS OF NEGATIVE HALLUCINATIONS OF
HYPNOTIC AND 0-3 SIMULATOR SUBJECTS

Condition	Hypnotic Ss	0-3 simu- lators
Report negative hallucination	2	7
Deny negative hallucination	14	7

Note.—Fisher's exact probability, $p < .05$.

demand to negatively hallucinate, whereas only one-eighth of the hypnotic Ss reported negative hallucinations. Assuming for the moment that compliance is the sole basis for these hallucinatory reports, the 0-3 simulators are significantly more compliant than the hypnotic Ss in a demand context in which report compliance is very probable.

Table 2 indicates that much the same kind of results occurred for reports of positive hallucinations on Displays III and IV. Although the data here did not reach significance, it is clear that more simulator than hypnotic Ss tended to report positive hallucinations in a context where such reports were implicitly demanded by the suggestions. Again, it seems that report compliance in the context of implicit demands to report positive hallucinations is more characteristic of simulating than of hypnotic Ss.

Tables 1 and 2 do not include data from the seven highly susceptible simulator Ss. Four of these Ss reported negative hallucinations to Displays I and II. The same four Ss reported positive hallucinations on Displays III and IV.

It was expected that the countersuggestion for hypnotic Ss to remember color chips that had previously been negatively hallucinated would reveal the extent to which registration of the color chips at some level had in fact occurred. The fact that only two hypnotic Ss initially verbalized negative hallucinations

TABLE 2

REPORTS OF POSITIVE HALLUCINATIONS OF HYPNOTIC
AND 0-3 SIMULATOR SUBJECTS

Condition	Hypnotic Ss	0-3 simu- lators
Report positive hallucination	6	9
Deny positive hallucination	10	5

Note.—Fisher's exact probability, ns .

TABLE 3

REPORTS OF POSITIVE HALLUCINATIONS OF HYPNOTIC
AND 0-3 SIMULATOR SUBJECTS DURING POSTEX-
PERIMENTAL INQUIRY

Condition	Hypnotic Ss	0-3 simu- lators
Report positive hallucination	7	0
Deny positive hallucination	9	14

Note.—Fisher's exact probability. $p < .01$.

makes it difficult to interpret the fact that half (i.e., one) of them accurately remembered the negatively hallucinated chip. The countersuggestive manipulation promises to be of some interest, however, and a successful replication of this aspect of the experiment on a larger N would have theoretical significance concerning the mechanisms of hallucinatory activity.

Reports of Subjects during PEI

So far, the authors have been concerned with verbal reports by Ss evoked prior to the prearranged signal of "Now the experiment is over." At this signal, the demand context changed. The S was to provide an *honest* report instead of a correct report which satisfied the suggestion-demand to hallucinate. As mentioned earlier, Bowers (1967) showed that a similar alteration did have a significant effect on Ss' reports of hallucinations. Although the PEI context is not literally demand free, the nature of the demands on S are, psychologically speaking, quite different. The constraints on Ss' reports are no longer to comply with the perceived purpose of the experiment, but rather to report honestly their actual experience. Whether or not one succeeds in altering the demand characteristics of the experimental situation should be apparent in part by the experiential report of the 0-3 simulators. There is no basis for believing that these Ss should experience hypnotically induced hallucinations, and hallucinatory reports made by these Ss would suggest that residual demands to comply were still operative. The use of unsusceptible Ss as controls thus helps to assess the success with which the demand context is truly altered. If in response to demands for honesty, 0-3 simulators report no hallucinatory ex-

perience, then the demand context of the PEI can reasonably be assumed to have been altered in such a way as to produce honest instead of correct reports. If one can demonstrate the absence of situational demands to report in correct ways, then PEI hallucinatory reports by hypnotic Ss take on high credibility.

The data gathered in the PEI showed a dramatic change by the 0-3 simulators, in the predicted direction. The most revealing data concerned Ss' responses to Question 5 which asked whether or not S was in the condition in which color chips were present on Displays III and IV. It will be recalled that confirming the presence of chips to Displays III and IV constituted a positive hallucination. As Table 3 indicates, not one of the 0-3 simulator Ss stated that he was in the condition in which color chips appeared on Displays III and IV; seven hypnotic Ss said they were. The fact that no simulator Ss reported color chips on Displays III and IV suggests that the situational demands for an hallucinatory report were successfully nullified. In light of this finding, there is correspondingly less chance that hypnotic Ss were responding to situational demands during the PEI to report hallucinations. Therefore, the hallucinatory reports, given by hypnotic Ss in this context, have some credibility claim.

Although the seven hypnotic Ss in question were alike in stating that color chips were present on Displays III and IV, the actual nature and vividness of the hallucinated color chips were quite variable. In fact, as the reader may have noticed, one hypnotic S who did not report color chips on Displays III and IV during the experiment reversed himself during the PEI. The remaining six Ss were the same ones who had previously reported positive hallucinations to Displays III and IV. One of these Ss reported a yellow triangle in the upper section of Display III and nothing on Display IV. Of course yellow was not a color that was used in this experiment. The two above Ss and four of the remaining five Ss who reported positive hallucinations responded to Question 6 that the (positively hallucinated) chips on Displays III and IV seemed less real than those remembered on

Displays I and II; the remaining *S* claimed the (hallucinated) chips on Displays III and IV were more real. To the seventh question, four *Ss* stated that the color chips on Displays III and IV appeared less real than the chips which had appeared on the original Displays A and B. One *S* stated that the chips on Displays III and IV were equally as real as those appearing on Displays A and B, and another *S* said the chips on Displays III and IV were more real than those on A and B. The last *S* (the one who had not positively hallucinated in the presence of the Displays III and IV) said that a red chip seen on Display III was less real, and a green chip on Display IV was equally as real as those chips appearing on Displays A and B.

It is clear that the positive hallucinations were seldom as vivid as actual color chips. However, the hallucinations were all sufficiently convincing to warrant *Ss*' acknowledgment of their external reality as a subjectively more valid option than would be the denial of their experiential reality altogether. Indeed, three out of the seven hypnotic hallucinating *Ss* stated in response to Question 10 that it would "now seem more incredible" that there were no chips on Displays III and IV than that there were no color chips on Displays I and II after all.

Although *Ss*' reports to Question 5 and subsequent questions were probably the most interesting aspect of the present findings, *Ss*' reports to Question 3 were also quite consistent with the data and arguments previously presented. This question asked whether the color chips on Displays I and II now seem less real, equally as real, or more real than those chips that had appeared on Dis-

TABLE 4

RESPONSES OF HYPNOTIC AND 0-3 SIMULATOR SUBJECTS TO QUESTION 3 OF THE POSTEXPERIMENTAL INQUIRY

Condition	Hypnotic <i>Ss</i>	0-3 simu- lators
Displays I & II chips more or equally as real as Displays A & B	10	14
Displays I & II chips less real than Displays A & B	6	0

Note.—Fisher's exact probability, $p < .02$.

TABLE 5

CHANGE IN REPORTS DURING THE POSTEXPERIMENTAL INQUIRY OF SUBJECTS WHO REPORTED POSITIVE HALLUCINATIONS DURING THE EXPERIMENT

Condition	Hypnotic <i>Ss</i>	0-3 simu- lators
No change in report from one context to another	6	0
Change in report from one context to another	0	9

Note.—Fisher's exact probability, $p < .001$.

plays A and B. All 0-3 simulators said they appeared equally as real, but hypnotic *Ss* were again more variable. Two hypnotic *Ss* said color chips on Displays I and II now seemed more real than the chips on Displays A and B; eight *Ss* said they were equally as real; and six *Ss* said they were less real. If the more real category is collapsed into the category equally as real, then the result, as seen in Table 4, can be tested by the usual Fisher's exact probability formula. These data seem to indicate that significantly more hypnotic than 0-3 simulator *Ss* experienced *some* perceptual distortion on Displays I and II, despite the fact that significantly more 0-3 simulator *Ss* originally reported negative hallucinations to Displays I and II (see Table 1).

A final indication of the extent to which the report context is vital for assessing the experiential reality of visual hallucinations can be observed in Table 5. Originally in the experiment, 6 hypnotic and 9 simulator *Ss* reported positive hallucinations. Of these 15 *Ss*, 6 maintained their original report of hallucinations during the PEI. Table 5 shows that these 6 were the hypnotic *Ss*. Thus, one sees that 0-3 simulators who reported positive hallucinations, reported them only in a context where such reports were implicitly demanded by the suggestions to do so. The hypnotic *Ss* who reported positive hallucination, however, did so *both* in a context where they were implicitly demanded *and* in a context relatively free of such demands.⁷

⁷ The fact that highly susceptible *Ss* who are told to fake hypnosis might inadvertently, and unwittingly, become hypnotized was the reason for employing unsusceptible *Ss* as controls in this experiment. Obviously, *S* confounding does occur in such a design, and this fact can vitiate the generality of the findings. It was in cognizance of this fact that

DISCUSSION

The two main findings of this experiment are: (a) hypnotic Ss who reported positive hallucinations continued to acknowledge them under the same PEI report conditions in which simulating Ss did not, and (b) significantly fewer hypnotic than simulator Ss acknowledged negative hallucinations under experimental report conditions, conditions in which these hallucinations were implicitly demanded by specific hypnotic suggestions.

If it were not for this second finding, the first finding would be a less convincing demonstration that some Ss actually experienced positive hallucinations. One might argue for instance that the PEI was perceived differently by simulating and by hypnotic Ss in ways that could account for differences in the hallucinatory reports. The simulators were, after all, explicitly directed to stop faking and to be honest during the PEI. The hypnotic Ss, on the other hand, were never told to simulate (i.e., lie) in the first place, so that any directive to be honest during the PEI might not have had the same significance that it had for simulating Ss. It is conceivable, therefore, that the hypnotic Ss were merely complying with some unwanted and cryptic demands, inherent in the PEI, to report positive hallucination.

It must be recalled, however, that only 2 out of 16 hypnotic Ss, as compared with 7 out of 14 simulator Ss, reported negative hallucinations *during* the experiment when such reports were implicitly demanded by the suggestions to negatively hallucinate. Compliance to regnant situational demands certainly

a small number of susceptible simulators were included in this experiment. With the exception of one S, the behavior of Ss in this group paralleled quite closely that of the O-3 simulators. For example, three out of the four susceptible simulators who reported positive hallucinations during the experiment later denied them in the PEI. Interpretations based on the data of the O-3 simulators do not, therefore seem to be misleading. The fact that susceptible simulators did tend to behave like O-3 simulators suggests (a) that susceptible Ss can resist hypnotic suggestions and (b) that the discrepancy in the number of hours that hypnotic and O-3 simulators had been seen for training was not a factor of overwhelming importance in the comparison of the data.

does not explain the verbal behavior of hypnotic Ss during the experiment. Indeed, it seems eminently reasonable to accept as credible, the verbal reports of hypnotic Ss indicating the *absence* of negative hallucinations when these reports occur in an experimental context implicitly demanding such hallucinations. It seems no less reasonable to assume the credibility of positive hallucinations tendered during the PEI; for the PEI constitutes a report context which is quite probably less demanding of verbal compliance than is the experimental situation itself. Inasmuch as the hypnotic Ss reported relatively few (negative) hallucinations in a context where such reports were implicitly demanded, and comparatively more (positive) hallucinations in a context relatively free of demands to do so, it would appear that hallucinatory reports of hypnotic Ss are more credible than they are compliant to situational demands.

If, on the basis of the evidence and arguments here proffered, one can assume that hypnotic Ss tend to give credible reports, then one can also assume that their reports are very important data. In fact, Ss' verbal reports about their hypnotic experiences can then be seen to constitute evidence by which to ascertain whether or not a person was hypnotized. Having such data helps us to wiggle free of the tautology that Barber (1964) has pointed to wherein "a person is said to respond to suggestions because he is in hypnosis and he is said to be in hypnosis because he responds to suggestions [p. 839]." According to the view espoused here, an S's testimony about how he experienced his behavior is critical in "cracking" the tautology, because it constitutes independent evidence about S's actual experiential state.

The work of Hilgard and Tart (1966; Tart & Hilgard, 1966) is very consonant with the above line of reasoning. They elicited periodic state reports from Ss which served as criteria of hypnotic depth that were "relatively independent of the responses to various tests of suggestion [Hilgard & Tart, 1966, p. 203]." As it turns out, these state reports of hypnotic depth correlated with suggestibility, and they did so even in Ss who had *not* received formal hypnotic induction. The use of their

experimental technique thus allows a conclusion that some nonhypnotic control Ss in fact become hypnotized—rather a turnabout from a more familiar conclusion that hypnotic Ss are merely task motivated.

In conclusion, the present study supports other recent reappraisals suggesting the inadequacy of the motivated compliance view of hypnosis (Bowers, 1966; Evans, 1966; Evans & Orne, 1965). It has been argued, here and elsewhere (Orne, 1966), that the subjective characteristics of hypnotic behavior are of paramount importance in any attempt to understand the phenomena. Consequently, the experiential reports of hypnotized Ss become very significant data, especially when there is new cause to regard them as credible. It seems to the authors that it has been an unwillingness of some investigators to take seriously the subjective reports of hypnotic Ss that has opened the gates of uncritical skepticism concerning hypnotic phenomena. Perhaps the data and arguments offered herein will persuade other workers that hypnotic Ss have something important to say.

REFERENCES

- BARBER, T. X. "Hypnosis" as a causal variable in present-day psychology: A critical analysis. *Psychological Reports*, 1964, 14, 839-842.
- BARBER, T. X., & CALVERLEY, D. S. "Hypnotic behavior" as a function of task motivation. *Journal of Psychology*, 1962, 54, 363-389.
- BARBER, T. X., & CALVERLEY, D. S. An experimental study of "hypnotic" (auditory and visual) hallucinations. *Journal of Abnormal and Social Psychology*, 1964, 68, 13-20.
- BEM, D. J. An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, 1965, 1, 199-218.
- BOWERS, K. S. Hypnotic behavior: The differentiation of trance and demand characteristic variables. *Journal of Abnormal Psychology*, 1966, 71, 42-51.
- BOWERS, K. S. The effects of demands for honesty on reports of visual and auditory hallucinations. *International Journal of Clinical and Experimental Hypnosis*, 1967, 15, 31-36.
- EDMONSTON, W. E. Stimulus-response theory of hypnosis. In J. E. Gordon (Ed.), *Handbook of clinical and experimental hypnosis*. New York: Macmillan, 1967.
- EVANS, F. J. The case of the disappearing hypnotist. Paper presented at the meeting of the American Psychological Association, New York, September 1966.
- EVANS, F. J., & ORNE, M. T. Motivation, performance, and hypnosis. *The International Journal of Clinical and Experimental Hypnosis*, 1965, 13, 103-116.
- HILGARD, E. R., & TART, C. T. Responsiveness to suggestions following waking and imagination instructions and following induction of hypnosis. *Journal of Abnormal Psychology*, 1966, 71, 196-208.
- HULL, C. L. *Hypnosis and suggestibility: An experimental approach*. New York: Appleton-Century, 1933.
- JONES, E. E., & DAVIS, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. II. New York: Academic Press, 1965.
- ORNE, M. T. Hypnotically induced hallucinations. In L. J. West (Ed.), *Hallucinations*. New York: Grune & Stratton, 1962.
- ORNE, M. T. Hypnosis, motivation and compliance. *The American Journal of Psychiatry*, 1966, 122, 721-726.
- SHOR, R. E., & ORNE, E. C. *Harvard Group Scale of Hypnotic Susceptibility*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- TART, C. T., & HILGARD, E. R. Responsiveness to suggestions under "hypnosis" and "waking-imagination" conditions: A methodological observation. *The International Journal of Clinical and Experimental Hypnosis*, 1966, 14, 247-255.
- WEITZENHOFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto: Consulting Psychologists, 1962.

(Received April 1, 1968)

CYCLIC VARIATION IN STUTTERING:

COMMENT ON TAYLOR AND TAYLOR'S "TEST OF PREDICTIONS FROM THE CONFLICT HYPOTHESIS OF STUTTERING"

JOSEPH SHEEHAN

University of California, Los Angeles

Although stuttering behavior may show some constancy over short time sequences as Taylor and Taylor have shown, larger time segments do show a cyclic variation. Brief readings of experimental passages do not always encompass the range of cyclic occurrence. Data from Quarrington and from the author are cited in support of the approach-avoidance conflict theory of stuttering, particularly the fear-reduction hypothesis: viz., that the occurrence of stuttering behavior reduces the fear that elicited it.

An intriguing feature of the mystery of stuttering is that most stutterers are able to speak most of their words fluently, as Johnson, Brown, Curtis, Edney, and Keaster (1967) have shown. A wavelike, periodic fluctuation in frequency of stuttering has added to the mystery.

During the course of exposition of a conflict theory of stuttering, Sheehan (1958) posed two questions related to stuttering as a momentary blocking of speech: (a) What makes the stutterer stop? (b) Once stopped, what enables him to continue? To answer these questions, Sheehan proposed two major hypotheses: (a) the conflict hypothesis—the stutterer stops whenever conflict approach and avoidance tendencies reach an equilibrium; (b) the fear-reduction hypotheses—the occurrence of stuttering behavior reduces the fear that elicits it. Experimental evidence for both hypotheses was covered in connection with the theory presentation and need not be reviewed here.

As a corollary or outgrowth of the fear-reduction hypothesis, the theory suggested that the wavelike variation noted in stuttering might be explained as follows: During disfluent periods, stuttering reduces fear sufficiently to permit fluency, so that there is a buildup of fluency due to actual occurrence of stuttering behavior. During fluent periods the fear is not dissipated and builds up, for the fluency itself sets up a role expectation

that makes stuttering more probable. In this way stuttering leads to fluency and fluency leads to stuttering. Hence the wavelike, cyclic pattern of stuttering occurrence.

Recently in this journal, Taylor and Taylor (1967) offered certain data as a test of "the Sheehan-Quarrington conflict hypothesis." More precisely, they should have referred to the fear-reduction hypothesis, just summarized in the preceding paragraph. I. K. Taylor (1966) had had nine stutterers read a 500-word passage three times in succession. Further analysis of these data showed "no tendency either for periodicity or for clustering of the stuttering events [Taylor & Taylor, 1967]." Apparently, from these data the Taylors felt that they had disproved the existence of cyclic variation in stuttering. At least, on the basis of this finding they questioned the whole approach-avoidance conflict interpretation of stuttering (Sheehan, 1953).

The ambiguity here concerns the duration of the phases of the cyclic variations of stuttering, that is, the length of the waves.

From clinical observations and from Quarrington's² data it may be said that there are large waves and ripples, and ripples within larger waves. The crests and troughs may vary widely from one individual to another, and within the same individual at different times. But the cyclic characteristic of stuttering is not confined to a narrow time segment such as that studied by Taylor and Taylor. The pattern of variation may be second to second, minute to minute, hour to hour, day

¹ Requests for reprints should be sent to the author, Department of Psychology, University of California, Los Angeles, California 90024.

² B. Quarrington, personal communication, 1968.

to day, week to week, month to month, and even year to year. Stuttering cycles may also vary according to the life stage of the individual stutterer (Conway & Quarrington, 1963; Quarrington, 1965).

In support of this pattern of variation, Quarrington has recently gathered data on 39 young stuttering children, mean age 4 yr., 4 mo., observed over a 1-yr. period. In 32 of these stuttering persisted over the 1-yr. period, during which their mothers supplied daily ratings of the frequency of stuttering. When submitted to Fisher's harmonic analysis, the resulting data curves revealed the presence of significant periodicities. Fifty-nine percent of the persistent stutterers showed statistically significant periodic components in the cycle length range of 2-6 mo. There was no evidence of shorter periodicities, nor was there evidence of any weekly patterning of stuttering frequency.

Since stuttering behavior is notoriously dependent on situational variables, Taylor and Taylor should hardly have expected to encompass the range of wavelike variation in stuttering within the minutes it takes to read an experimental passage. Yet that is apparently what they assumed.

The Taylor and Taylor data are interesting but not astonishing in showing that stuttering may occur with some constancy over short time sequences. But every study of the adaptation effect, involving gradual reduction of stuttering frequency over successive readings of the same passage, has shown a similar constancy. The time span it takes to read

passages during an experimental session is really quite short: 15 min., or 30, but seldom over an hour even for the most severe stutterer. These are not the predominant orders of time referred to in describing stuttering behavior as cyclic.

Neither the Sheehan conflict theory nor the case for cyclic variation in stuttering rest primarily upon the minute time segments that Taylor and Taylor apparently expected, but naturally enough did not find. The waves in stuttering are those of a big ocean, not of a small puddle.

REFERENCES

- CONWAY, J. K., & QUARRINGTON, B. Positional effects in the stuttering of contextually organized verbal material. *Journal of Abnormal and Social Psychology*, 1963, 67, 299-303.
- QUARRINGTON, B. Stuttering as a function of the information value and sentence position of words. *Journal of Abnormal Psychology*, 1965, 70, 221-224.
- JOHNSON, W., BROWN, S., CURTIS, J., EDNEY, C., & KEASTER, J. *Speech handicapped school children*. (3rd ed.) New York: Harper & Row, 1967.
- SHEEHAN, J. Theory and treatment of stuttering as an approach-avoidance conflict, *Journal of Psychology*, 1953, 36, 27-49.
- SHEEHAN, J. G. Conflict theory of stuttering. In J. Eisenstein (Ed.), *Stuttering: A symposium*. Harper, 1958.
- TAYLOR, I. K. The properties of stuttered words. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 112-118.
- TAYLOR, I. K., & TAYLOR, M. M. Test of predictions from the conflict hypothesis of stuttering. *Journal of Abnormal Psychology*, 1967, 72, 431-433.

(Received April 8, 1968)

VERBAL OPERANT CONDITIONING OF HOSPITALIZED PSYCHIATRIC PATIENTS

WILLIAM DRENNEN,¹ WILLIAM GALLMAN, AND GENE SAUSSER²

University of South Carolina

Two experiments were conducted. In Experiment I, 81 Ss were divided into nine subgroups according to age and length of hospitalization. Plural nouns were reinforced by positive verbal feedback. Results were (a) least chronic Ss increased their operant rates; (b) most chronic Ss did not; (c) there was significant interaction with respect to age and chronicity. In Experiment II, 60 schizophrenics Ss were divided into six subgroups with respect to length of hospitalization and mode of reinforcement (i.e., verbal, primary, combination). Acute-type schizophrenics responded positively to all types of reinforcement. Chronics responded only to primary reinforcement and to the combination procedure. Mode of reinforcement was a significant variable affecting operant rate.

Verbal operant conditioning with psychiatric patients residing in a hospital or institutional setting has received increasing attention in the past few years. Particular attention has been devoted to research with patients with a diagnosis of schizophrenia even though the reliability of diagnosis is seriously open to question (Mehlman, 1952; Schmidt & Fonda, 1956). The literature dealing with verbal conditioning of normals and psychiatric patients has demonstrated consistently that normals condition much more readily, at least when verbal reinforcement is used as feedback (Buss & Lang, 1965).

Baron (1966), in attempting to explain these findings, suggested that schizophrenics have had a negative premorbid social reinforcement history for a variety of behaviors. Baron suggested that schizophrenics evaluate positive verbal-social reinforcement as inappropriate and thus fail to respond in a positive manner. However, some patients diagnosed as schizophrenic are apparently more responsive than others and condition more readily (Johannsen & Campbell, 1964). In addition, patients who leave the hospital sooner (i.e., within 180 days) condition more readily, apparently because they are of the

acute rather than chronic type (Salzinger & Portnoy, 1964). In related research, True (1962) found that reactive schizophrenics conditioned more rapidly than those classified as process. It has further been demonstrated by a number of research studies that schizophrenics respond or learn more under the conditions of verbal censure than verbal reward.

Schizophrenic patients seem to react to primary or nonverbal reinforcement more consistently than to verbal reinforcement. Kilberg (1962) found that nonverbal rewards in the form of cigarettes *did not* lead to significant differences in verbal conditioning of normals and schizophrenics. The more general operant conditioning procedures investigated by Ayllon and Haughton (1962, 1964) also supports this hypothesis. Further classification and exploration in this area of behavioral modification appears warranted, especially with regard to variable interaction, extinction, generalization effects, and combinational procedures. This type of experimentation, in addition to providing information with regard to principles of human learning, has practical implications related to more effective treatment of institutionalized psychiatric patients. It is not enough to say that chronic psychiatric patients and/or schizophrenics are "different" and "unresponsive," but also to circumvent these difficulties by devising effective procedures to increase adaptive behaviors and extinguish maladaptive behaviors.

¹ Requests for reprints should be sent to William Drennen, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

² The authors are grateful to the patients and staff of the South Carolina State Hospital for their cooperation in this study.

TABLE 1

SUBJECTS WITH RESPECT TO CHRONICITY: AGE RANGE, NUMBER OF HOSPITALIZATIONS, AND DIAGNOSES

Chronicity	Age ranges (no. Ss)			Mean number of hospitalizations	Diagnoses* (no. Ss)			
	26-42	35-43	44-52		1	2	3	4
6 mo.	9	9	9	1.01	17	3	0	7
1-2 yr.	9	9	9	1.85	20	1	1	5
3-5 yr.	9	9	9	1.93	21	0	2	4

* 1 = schizophrenia; 2 = psychoneurosis; 3 = brain syndrome; 4 = personality disorder.

The purposes of the two experiments conducted were to ascertain the changes in verbal operant response rate as a function of age, length of hospitalization (i.e., chronicity level), interaction of these variables, and finally the effects of mode of reinforcement (i.e., verbal, primary, or a combination of the two of these). Patient awareness was also assessed as a variable affecting learning and was ascertained in the manner of Denike (1964).

EXPERIMENT I

Method

Subjects. The Ss consisted of 81 patients residing in a state mental hospital. All Ss were male with at least a fifth-grade reading level as ascertained by use of the Wide Range Achievement Test (Jastak & Bijou, 1946). The Ss were divided into three major age and chronicity groups as determined by total length of time spent in the hospital without regard to number of hospitalizations. Table 1 presents a summary of S group selection.

Procedure. The verbal operant conditioning procedure was similar to that used by DeNike (1964). Human nouns constituted the critical response class reinforced verbally by E saying "good" or "fine." Each S was required to say a total of 200 words singly with E facing S. The S was instructed to not repeat words. The conditioning task of 200 words was divided into eight trial blocks of 25 words each for purposes of statistical analysis. No reinforcement was given during the first trial block in order to establish operant levels of the Ss. To ascertain awareness, after each trial block Ss were given 1 min. to express verbally any thoughts they might have about the experiment. After the entire procedure, Ss were asked a series of questions devised to assess awareness of reinforcement contingency of the human noun concept or of any related concept (e.g., Did you notice that I said anything? What kind of words did you say most? Why?) To be classified aware an S had to say either that E said "good" or "fine" when he gave human nouns, words about people, or illustrated this concept by example.

Results

Individual scores were obtained by subtracting the number of human noun responses in the operant base-line block from the number emitted in the last trial block. Because the within-group variance was heterogeneous, analyses of the data were carried out both with raw and transformed scores. (Square-root transformations were necessary to obtain adequate within-group homogeneity of variance.) A factorial analysis revealed that between-group differences were statistically significant with respect to the chronicity of Ss ($p < .01$), but not with respect to age. The interaction effect Age \times Chronicity Level was highly significant ($p < .01$). Inspection of the data suggested that the interaction effect Age \times Chronicity occurred most markedly in the middle-chronicity-level group where older Ss responded or learned more efficiently than younger Ss. Group learning differences generally were, as expected, in favor of the acute-type patients. That is, Ss with shorter hospitalization learned relatively more in terms of increased operant rate.

Only 12 of the 81 Ss expressed any form of awareness of reinforcement contingency. Six of these Ss were of the acute type, 2 were moderately chronic, and 4 were chronic-type patients. In the acute group of patients between-group differences were not significant ($p > .10$) with respect to Ss' awareness. In the other groups (i.e., moderately chronic and chronic) there were not enough aware Ss to test for significance of between-group differences.

EXPERIMENT II

Method

Subjects. Sixty male patients between the age of 20 and 35 yr. who met the following qualifications

TABLE 2
REINFORCEMENT PROCEDURES FOR ACUTE AND
CHRONIC GROUPS

Subjects	20 trials	70 trials	70 trials	70 trials
Acute				
A ₁	Operant	PRVR	PRVR	VR
A ₂	Operant	PR	PR	VR
A ₃	Operant	VR	VR	VR
Chronic				
C ₁	Operant	PRVR	PRVR	VR
C ₂	Operant	PR	PR	VR
C ₃	Operant	VR	VR	VR

Note.—A and C refer to acute or chronic, respectively. VR, PR, and VRPR refer to method of presentation of reinforcement with VR being verbal reinforcement, PR primary reinforcement, and VRPR a combination of the two. $N = 10$ in each subgroup.

were selected: (a) at least a seventh-grade education; (b) at least a fifth-grade word pronunciation level (Wide Range Achievement Test, Jastak & Bijou, 1946); (c) either delusional, hallucinating, or both; and (d) independently diagnosed as schizophrenic by a hospital staff psychiatrist.

In addition, 30 of the Ss selected had been in the hospital less than 6 mo.; 30 had been hospitalized for more than 3 yr. These two major groups were classified as either acute (less than 6-mo. hospitalization) or chronic (more than 3-yr. hospitalization). Each of the major groups was subdivided into three treatment subgroups of 10 Ss each. Assignment to the treatment subgroups was on a random basis.

Table 2 summarizes the experimental paradigm. The verbal operant conditioning procedure was again similar to that used by DeNike (1964). All Ss were instructed to say words singly, without repeating words. Human nouns were reinforced after each S had said the first 20 words without interruption. Each S said a total of 230 words.

As indicated in Table 2, acute and chronic Ss in each of the three treatment subgroups received a different mode of reinforcement. The type of primary reinforcement used was that of candy and cigarettes. Verbal reinforcement consisted of the E saying "good" or "fine." The combination procedure involved the simultaneous use of both of these types of feedback. It should be noted that in the last block of 70 trials, all Ss received only verbal reinforcement.

The Ss were given 1 min. to verbalize thoughts about the experiment at trial-block intervals. At the end of all trials, Ss were asked a series of questions as in Experiment I to elicit any evidence of awareness of reinforcement contingency.

Results

Subgroup means are presented in Table 3. Initial operant rates did not differ significantly ($p > .10$). Scores used in further statistical comparisons were determined by subtracting each of the initial Ss operant base-line rates from the operant in the trial block analyzed. Square-root transformations were found to be necessary in order to gain homogeneity of within-group variance.

A 2×3 factorial analysis was performed to test for between-group differences and interaction effects on the data of the second block of 70 trials. Mode of reinforcement resulted in significant between-group differences ($p < .05$). Chronicity level and interaction effects were not significant. The same type analysis was made with respect to the last block of 70 trials with similar results. In fact, as illustrated in Table 3, very little change occurred in the last block of 70 trials in which all Ss experienced the same mode of reinforcement. Inspection of the data suggests that, as might be expected from previous research, the groups responded very much the same with the exception of the chronic subgroup which received only verbal reinforcement on all trials. A series of t tests comparing each subgroup with each of the others in the last 70 trial block revealed that all groups were superior to the chronic group receiving only verbal reinforcement. However, the significance level of the acute verbal reinforcement/chronic verbal reinforcement comparison was not high enough to discount chance differ-

TABLE 3
SUBGROUP MEANS IN TRIAL BLOCKS

Group	Operant 20 trials	Block 1 70 trials	Block 2 70 trials	Block 3 70 trials
Acute verbal reinforcement	2.6	6.4	6.6	7.0
Acute primary reinforcement	1.9	4.5	9.9	8.2
Acute primary reinforcement-verbal reinforcement	1.8	8.1	10.1	11.6
Chronic verbal reinforcement	2.3	3.2	3.6	3.1
Chronic primary reinforcement	0.6	5.3	7.0	7.1
Chronic primary reinforcement-verbal reinforcement	1.5	8.3	7.9	6.7

ences. All other subgroups were clearly superior to the chronic-verbal-reinforcement group ($p < .05$). No other subgroup comparisons (i.e., excluding the chronic-verbal-reinforcement group) approached a level of statistical significance ($p > .10$). As indicated in Table 3, the means of the subgroups receiving primary reinforcement or the combination procedure were highly similar.

Only 10 of the 60 Ss participating in the study verbalized any type of reinforcement contingency, 5 in the acute group and 5 in the chronic group. The aware- and unaware-acute Ss exhibited very similar operant levels on the last block of 70 trials. However, chronic-aware Ss receiving primary reinforcement or a combination of primary and verbal reinforcement clearly were superior to their unaware peers ($p < .01$).

DISCUSSION

The results of Experiment I clearly suggest that chronic psychiatric patients, ignoring diagnosis, react very little in a positive way to verbal reinforcement as compared with acute patients. The chronic patient maintains the status quo in terms of his verbal operant behavior or may even decrease the frequency of the particular operant involved. Age itself does not lead necessarily to differences in positive responsiveness but does interact with level of chronicity, especially in the middle level of chronicity (i.e., 1-2 yr.).

The results of Experiment II clearly demonstrate that primary reinforcement or a combination of primary and verbal reinforcement was more effective in producing an increased operant in chronic schizophrenics than verbal reinforcement alone. However, there did not seem to be any clear differential effects of the mode of treatment with respect to acute schizophrenics. The combination of two modes of reinforcement as compared with primary reinforcement alone did not lead to differential effects with either the acute or chronic schizophrenics.

An important question raised by the results was whether the schizophrenic patient who increased his operant rate of primary reinforcement would then regress when the mode of reinforcement was changed to positive verbal feedback. In this regard, there was no

marked change in operant level when modes of reinforcement changed. However, more trials after the transition would be considered necessary to adequately test out this possibility. The effects of combination procedures should be more thoroughly investigated in further research in view of the important treatment implications as well as basic research in learning.

Only about 21% of the patients verbalized any type of reinforcement contingency. From both experiments there were 11 aware-acute versus 9 aware-chronic patients out of a total of 94 so classified. This is somewhat surprising in that one would expect the acute patients to be more "alert" as to what was taking place. Motivation on the part of the aware Ss was evaluated by interview. There was no clear relationship between what Ss said and what they did. Their answers were somewhat vague or contradictory. These difficulties of assessing motivation were apparently not encountered by DeNike (1964). However, in a recent unpublished study by Wagoner (1967) using college-type Ss, difficulties of assessing motivation to receive a verbal reward were encountered. When some Ss solved the concept, they tested out other hypotheses, or at times decided not to respond in the same way. Others did the opposite of what they said they were motivated to do. It is possible also that Ss may reach a plateau of positive response, as has been suggested by Baron (1966). For the time being at least, one is led to conclude that an S's expressed awareness may be neither necessary nor sufficient as a factor influencing an increased operant rate. What a person says and does is not always consistent. This does not argue against S's assessed awareness, however, as being an important variable involved in any type of learning situation.

REFERENCES

- AYLLON, T., & HAUGHTON, E. The control of behavior of schizophrenic patients by food. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 342-352.
- AYLLON, T., & HAUGHTON, E. Modification of symptomatic verbal behavior of mental patients. *Behavior Research Therapy*, 1964, 2, 87-97.
- BARON, R. M. Social reinforcement effects as a function of social reinforcement. *Psychological Review*, 1966, 73, 527-539.

- BUSS, A. H., & LANG, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, concept attainment. *Journal of Abnormal Psychology*, 1965, 70, 2-24.
- DENIKE, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology*, 1964, 68, 521-529.
- JASTAK, J., & BIRCH, S. *Wide Range Achievement Test*. (1st ed.) Wilmington: Guidance Associates, 1946.
- JOHANSEN, W. F., & CAMPBELL, J. Verbal conditioning in chronic schizophrenia: Effects of reinforcement class and social responsiveness. *Psychological Report*, 1964, 14, 567-672.
- KILBERG, J. The differential effects of nonverbal and verbal behavior in the modification of verbal behavior of schizophrenic and normal subjects. Unpublished doctoral dissertation, Columbia University, 1962.
- MEHLMAN, B. The reliability of psychiatric diagnosis. *Journal of Abnormal and Social Psychology*, 1952, 47, 577-578.
- SALZINGER, K., & PORTNEY, S. Verbal conditioning in interviews: Application to chronic schizophrenics and relationship to prognosis for acute schizophrenics. *Journal of Psychiatric Research*, 1964, 2, 1-9.
- SCHMIDT, H. O., & FONDA, C. P. The reliability of psychiatric diagnosis: A new look. *Journal of Abnormal and Social Psychology*, 1956, 52, 262-267.
- TRUE, J. E. The verbal conditioning of abstract responses in process and reactive schizophrenic patients. *Dissertation Abstracts*, 1962, 23, 711-712.
- WAGONER, P. Response sets and awareness in verbal conditioning. Unpublished master's thesis, University of South Carolina, 1967.

(Received April 26, 1968)

PERCEPTUAL STYLES AND THEIR CORRELATES AMONG SCHIZOPHRENIC PATIENTS

CARMI SCHOOLER AND JULIAN SILVERMAN¹

National Institute of Mental Health, Bethesda, Maryland

The experimental procedures on which an earlier tridimensional model of schizophrenic perceptual functioning was based were replicated, other potentially relevant procedures included, and the resultant relationships delineated and tested through multivariate statistical techniques. Factor analysis supported the existence of the three dimensions of the original model—stimulus intensity control, scanning control, and field articulation. Through cluster analysis there emerged groups of Ss whose differing perceptual styles paralleled the previously reported pattern of differences between paranoid and nonparanoid schizophrenics. Other results, however, made several major qualifications of the original theory necessary: (a) perceptual inefficiency and anchoring emerge as additional meaningful dimensions of perceptual functioning; (b) behavioral guardedness replaces coherent paranoid delusions as a significant correlate of scanning; and (c) stimulus augmentation is related to strong internal affect only among acute patients; among chronic patients the relationship is reversed.

Through the examination of the relationships between styles of cognitive functioning and so-called "psychophysiological defense mechanisms," one of the authors has outlined a tridimensional model of perception that can be applied to the study of both psychopathological and normal groups. A full description of the model can be found in Silverman (1964b, 1967c, 1968). In brief, the dimensions involved are stimulus intensity control, scanning control, and field articulation.

Stimulus intensity control refers to the intensity with which perceptions are experienced; it might also be considered the intensity dimension of attention. In sensory response studies conducted with low or ordinary ranges of stimulation, hypersensitive individuals experience stimuli intensely; they react strongly to such input. Under conditions of high-intensity stimulation, or of sudden or otherwise novel stimulation, their responses

are attenuated. Such individuals are labeled reducers. Apparently because of their automatic tendency to attenuate high-intensity sensory input they are relatively tolerant of painful stimulation (Petrie, 1967; Silverman, 1964a). On kinesthetic figural aftereffect procedures (Petrie, 1967; Silverman, 1967c), reducers underestimate the size of a "test bar" after having rubbed an interpolated stimulation bar for a period of time. Those on the other end of this dimension, who are termed augmenters, show a notable intolerance for pain but a relative tolerance for reduced sensory input. They respond on the kinesthetic figural aftereffect procedures by overestimating the test bar after having rubbed the interpolated stimulation bar. Performance on these procedures can thus be used as indicators of position on the stimulus control dimension.

The second underlying dimension, scanning control, is concerned with the extensiveness with which individuals sample visual stimuli. Extensiveness of scanning is determined either directly from eye-movement recordings or inferred from procedures that are known to correlate with such recordings. The indirect method used most frequently is a particular form of size-estimation procedure (Gardner & Long, 1962). On this procedure minimal scanning of the visual field is associated with large overestimations of the size of visual stimuli.

¹The authors are extremely indebted to Frances Prestianni for her patience and diligence in gathering the perceptual data. They also wish to thank the staff of St. Elizabeths Hospital, Washington, D. C., where data were gathered, for their cooperation. The views expressed, of course, are those of the authors and do not necessarily reflect the point of view of the Hospital.

Requests for reprints should be sent to Carmi Schooler, Laboratory of Socioenvironmental Studies, National Institute of Mental Health, Bethesda, Maryland 20014.

Field articulation, the third underlying perceptual dimension, is inferred from tasks that necessitate attention to certain segments of a stimulus field and simultaneous inhibition of attention to other segments of the field. This dimension has been conceptualized as encompassing Witkin's concept of field dependence-independence (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954).

Studies of these response dimensions in heterogeneous schizophrenic populations have indicated that acute schizophrenics with good premorbid histories and coherently expressed delusional systems tend to show extensive scanning and augmentation of the intensity of stimulation. At the opposite pole are the nonparanoid patients with poor premorbid histories who tend to evidence relatively minimal scanning and a reduction of the experienced intensity of stimulation (Silverman, 1967c). At the time of the formulation of the tri-dimensional model clear-cut differences between schizophrenic subtypes had not been found on the field articulation dimension. Recent articles by Sugerma and Cancro (1968) and Silverman (1968) indicate the unusual complexity of this dimension when Ss studied are schizophrenics.

In the present study perceptual variables and psychological and personal history variables were examined with an aim toward evaluating the above conceptual framework in greater detail. The same experimental procedures on which the formulation was based were employed, and the resultant relationships were delineated and tested by means of multivariate statistical techniques. Other potentially relevant variables that had not been previously evaluated in this context were also included, both to extend the scope of the model and to evaluate possible alternatives to it.

The following perceptual procedures from which the dependent variables were derived were selected because they were central in developing the original model: visual size-estimation procedure (Silverman, 1964b); a test of kinesthetic figural aftereffects (Silverman, 1964a); a measure of "field independence," the rod-and-frame test (Witkin, 1965); and a measurement of category width (Pettigrew, 1958; Silverman, 1964c). In-

cluded because of their possible relevance were the Müller-Lyer Illusion Test (Gardner, 1961), the two-flash threshold procedure (Venables, 1963), and the Petrie small and large block stimulation procedures—a somewhat different and more elaborate measure of kinesthetic figural aftereffects than that used by Silverman (Petrie, 1967).

The independent variables were measures of characteristics of intellectual and psychosocial functioning and personal history that, on the basis of previous research, were expected to distinguish schizophrenic patients with apparently opposite perceptual styles: (a) acute paranoid schizophrenics with good premorbid histories and coherently expressed delusions, and (b) chronic nonparanoid schizophrenics with poor premorbid histories (Silverman, 1967c). A separate measure of each of these psychosocial and personality variables was included in order to assess the relative importance of each in predicting perceptual style. Selected on this basis were: length of hospitalization, paranoid versus nonparanoid subtype, Phillips Premorbid History Ratings (Phillips, 1953), and the separate items of the Wing Scale of Paranoid Behavior (Wing, 1961).

Other independent variables, effective in predicting the performance of schizophrenics in other studies, were included in an attempt to evaluate other aspects of abnormal behavior in relation to perceptual style. Central among these were measures of intellectual functioning: the Digit Span, Similarities, and Vocabulary subtests of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), and a clinical interview-based rating of adjustment (Rackow, Napoli, Klebanoff, & Schillinger, 1953). All of these measures are concerned with the degree of efficiency with which the schizophrenic patient deals with his environment. Personal history variables that were examined because of their possible effects on perceptual performance included age, age at first hospitalization, education, and social class of origin. Finally, two instruments—the Lorr Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr & Klett, 1965) and the Spohn Ward Behavior Scale (Schooler & Scarr, 1962)—were included in order to provide a fuller picture of each pa-

tient's present symptomatology and social behavior.

METHOD

Subjects

All Ss were white male hospitalized schizophrenics between 17 and 57 yr. of age who had no record of organic impairment. Acute and chronic patients of both paranoid and nonparanoid diagnostic subtypes were tested. All of the chronic patients had a history of more than 4 yr. of hospitalization and had been hospitalized more than 18 mo. since their most recent admission. Acute patients were those who were newly admitted to St. Elizabeths Hospital, Washington, D. C., who had a total hospitalization time of less than 2 yr., and who had not been hospitalized in any hospital more than 6 mo. during the present year.

Somewhat different criteria were used to select the different subgroups. Among the chronics an attempt was made to test all paranoid schizophrenic patients who had either (a) attended college or (b) not gone to high school. This criterion was included originally in an effort to evaluate the effects of education. Nonparanoid chronics were selected by choosing, for each paranoid used, the alphabetically next nonparanoid patient on the roster of the same hospital division who met the appropriate criteria and who was of the same educational level as the selected paranoid S. This educational criterion proved, however, too restrictive in terms of the number of patients available. Because of this, an equal number (4) of paranoid and nonparanoid patients who had attended high school were then randomly selected for inclusion in the sample. Of the 23 chronic paranoid patients so chosen, 2 refused and 1 was mute; of the chronic nonparanoid patients, 2 refused and 6 were mute. The final sample of chronic schizophrenics consisted of 20 paranoid and 15 nonparanoid Ss. The acute sample consisted of all patients admitted during a 3-mo. period who were diagnosed as schizophrenic and who met the appropriate criteria of hospitalization, age, race, and lack of organic impairment. Nineteen patients met this criteria. Two nonparanoid patients were unable to perform the tasks and 1 paranoid and 1 nonparanoid patient refused. The final sample of acute schizophrenics consisted of 9 nonparanoid and 6 paranoid Ss.

Procedure

Each S was seen on three occasions. During the first session he was interviewed by the senior author. On the basis of the interview he was evaluated on the IMPS (Lorr & Klett, 1965), the Wing Scale (Wing, 1961), and the Montrose Rating Scale (Rackow et al., 1953).²

² In a pilot study of reliability of a group of three trained raters, which included the interviewer (Schooler), the average interobserver correlation was greater than .90 for all but one of the IMPS factors (Motor Retardation = .74); the average intercorrelation

A battery of perceptual procedures and WAIS subtests were administered in the subsequent two testing sessions; each session lasted about 90 min. The intelligence tests were administered according to the description in the WAIS manual; the format for administering the perceptual procedures was essentially the same as that used in the previously cited perceptual studies. Procedures included in the second session were: (a) size-estimation procedure, (b) WAIS Vocabulary subtest, (c) two-flash threshold procedure, (d) Petrie large block stimulation procedure, and (e) Titchner Circles Illusion (Wapner & Werner, 1957). Included in the third session were: (a) WAIS Object Assembly Task, (b) Silverman kinesthetic figural aftereffect procedure, (c) Witkin rod-and-frame test, (d) Müller-Lyer illusion, (e) WAIS Digit Span, (f) Petrie small block stimulation procedure, and (g) Pettigrew category width test.

Perceptual Procedures

Titchner Circles Illusion. This illusion consisted of two inner circles, each of which was surrounded by a concentric ring of circles. The standard, always on the left, was composed of an inner circle, 16.5 mm. in diameter, and five surrounding circles, each with a diameter of 25 mm. The inner circle of the variable, always on the right, varied from 9.00 to 21.00 mm. in steps of .75 mm. The S was presented with a chart on which were mounted seventeen 4.5 in. × 7 in. cards on which a standard and a variable were drawn. The inner circle of the variable configuration was smallest at the top of the chart and increased in size with each succeeding card. The S was asked to judge for each of the 17 pairs whether the center circle on the right (variable) was smaller, equal, or larger than the center circle on the left (standard). On the first of two trials, S made his judgments beginning at the top of the chart and working down (D). On the second trial he began with the bottom pair and worked up (A). Performance scores were calculated in terms of the number of the card on which the S first responded to the illusion, if he responded at all. Card 11 was the one on which standard and variable were equal. Two scores were calculated for each S: (a) the average of the two responses (Titchner Mean) and (b) the difference between the score on the descending series and the score on the ascending series (Titchner D-A).

Müller-Lyer Illusion. The Müller-Lyer Illusion is an illusion of length: two objectively equal lines appear unequal, when at the end of one line small lines are drawn at acute angles to it, and at the end of the other line small lines are drawn obtusely to it

lation among observers across all of the factors being .91. A similarly high correlation with the ratings of other trained raters had previously been obtained with the Montrose Rating Scale. Although no direct interobserver reliability tests were made on the Wing Scale, the published evidence indicates that there is a high degree of agreement among professionally trained users.

(Gardner, 1961). The experimenter (*E*) gradually changes the length of the variable line until *S* says that it is the same length as the standard line. On half of the trials the variable line was started at a point distinctly smaller than the standard (ascending series), while in the other half of the trials it was started a point distinctly larger than the standard (descending series). The starting positions of the trials were varied in the order DAAD. Eight trials were presented. An *S*'s score on a given trial was the number of millimeters he erred from a correct judgment. Two scores were obtained: (a) the mean error based on all trials (Müller-Lyer mean) and (b) the mean error on descending trials minus the mean error on ascending trials (Müller-Lyer D-A).

Size-estimation procedure. Following Silverman's (1964c) procedure, *S* was asked to adjust a circular patch of light so that it appeared to be the same size as a hand-held disk. The light patch appeared on a double-ground glass screen 5 in. square that was mounted on a 9-in. square section of the end of a box. The 10-w. concentrated arc lamp (Sylvania C10 DC) was focused on the screen from inside the box to produce a smooth-edged circle of uniform brightness. The size of the circular patch projected on the glass screen was varied by having a circle cut in a piece of metal move back and forth on a spiral track as *S* turned a crank that was connected to the track.

Three disks were used: (a) a 10-gm. "neutral" gray disk 4.92 cm. in diameter, (b) a 65-gm. "neutral" gray disk 4.77 cm. in diameter, and (c) a 10-gm. disk covered with soft black wool 5.1 cm. in diameter. The *S* held each disk in his left hand, 15.5 in. away from his eyes while he adjusted the variable circle of light that was also 15.5 in. away from his eyes. Twelve judgments were made in four series of three judgments each, with a random order of disk size in each series. The order of ascending and descending judgments was DDAA. The *S* was instructed to adjust the circle of light so that it looked exactly like the size of the disk he was holding in his hand. At the end of each trial, *S*'s estimation was recorded, the light patch was turned off, and the mechanism reset at either ascending or descending starting position. No time limit was specified. Two scores were calculated: (a) the mean response on the 12 trials and (b) the difference between the mean score of the ascending trials and the mean score of the descending trials.

Kinesthetic figural aftereffect procedures. In the first procedure (Silverman, 1964a), four stimulus bars were mounted on saw horses at a convenient height for a standing *S*. The two test bars, between which the *S* stood as the procedure began, were a 1.5-in.-wide standard bar on his left and a comparison bar with twenty-six $\frac{1}{8}$ -in. gradations on his right. The wide end of the comparison bar, which was initially closest to *S*, measured 2.5 in.; the narrow end of the bar was 1 in. wide. The ninth gradation from the far end (narrow end) of the comparison bar was equal to the standard bar (1.5 in. wide). Directly behind the two test bars were two stimulation bars—

a .5 in.-wide bar on the left and a 3-in. bar on the right. All four bars were approximately 50 in. long. The blindfolded *S* first made four base-line judgments of where the graduated comparison bar felt just as wide as the standard bar. These four judgments were made from descending and ascending starting positions in DAAD order. The *S* then moved to the two stimulation bars and rubbed them simultaneously, three times per second for 30 sec., in time to a metronome. Following this interpolated stimulation he again judged the width of the 1.5-in. standard bar. This sequence was repeated until four judgments of the standard bar had been made following interpolation. The starting positions for these poststimulation judgments were the same as for the base-line series. Two separate scores were computed. The first score was obtained by calculating the difference in the chosen width of the comparison bar before and after each of the four interpolated trials (KFA total). An *S*'s score was the mean value for the four trials. The mean difference was also obtained for the ascending trials separately (KFA ascending).

The second figural aftereffects procedure employed the apparatus and technique used by Petrie (1967). The blindfold *S* made a series of six base-line judgments of the size of a standard block. Judgments were made by moving the left hand along a tapered block until a point was reached that felt equal in width to the standard block held in the right hand. Following this the *S* rubbed a stimulation block for a period of 90 sec., after which the initial procedure of judging the size of the standard block was repeated. The judgment procedure was again repeated following a 15-min. rest period. During the 45 min. preceding the base-line test and during the 15-min. rest period, *S* was required to avoid using his fingers in any way. This procedure was presented in both the second and third sessions, a 48-hr. interval separating them. In Session 2, in which the "large block stimulation test" was given, a 1 $\frac{1}{2}$ -in.-wide block was used as the standard and a 2 $\frac{1}{2}$ -in.-wide block for stimulation. In Session 3, during which the "small block stimulation test" was presented, a 2-in.-wide block was used as the standard and a 1-in.-wide block was used for stimulation. Two scores were computed from this experiment: (a) a difference score between the mean of the four judgments immediately preceding large block stimulation and the mean of the four judgments immediately following large block stimulation (Petrie large block stimulation), and (b) a difference score between the means of the four judgments immediately preceding and immediately following small block stimulation (Petrie small block stimulation).

Two-flash threshold procedure. The two-flash threshold was measured using a Grass Photo Stimulation Model PS2 with intensity set at 4. The lamp was located approximately 10 ft. from *S*. The *S* was asked to report whether he saw one or two flashes of light. Two trials were presented under ordinary room-illumination conditions. Each trial consisted of a different series of 10 two-flash intervals ranging in duration from 20 to 110 msec. The first series

started with very brief intervals between flashes, and the duration between pairs of flashes was gradually increased. The second series started with long intervals between flashes, and the duration between pairs of flashes was gradually decreased. The *S*'s two-flash threshold is the duration of the shortest interval at which two flashes are discriminated. Two scores were computed: (a) the average of the ascending and descending thresholds (two-flash mean) and (b) the difference between the descending and ascending two-flash thresholds (two-flash D-A).

Measure of category width procedure. The measure of category width consists of 20 multiple-choice items (Pettigrew, 1958). In each item an average value of a category is stated (e.g., the average speed of birds in flight) and *S* is asked to choose one of four alternatives for (a) the upper value of the category (the speed of the fastest bird) and for (b) the lower value of the category (the speed of the slowest bird). The range between these two values defines the limits of the category. The *S* was given the test booklet containing the items to be judged. Each item was also read to him before he made his choice. The scores used from these data were the Category Width II scores. This scoring used the 10 items that, according to Pettigrew's factor analysis of the 20-item questionnaire, were less influenced by quantitative reasoning ability (Silverman, 1964c).

Rod-and-frame test. The apparatus used for the Rod and Frame Test consists of a luminous square frame, pivoted at its center so that it can be tilted to left or right. Also pivoted at the center, but moving independently of the frame, is a luminous rod. During testing the room is darkened completely so that all the *S* can see is the frame and rod that are presented in various tilted positions. The *S* is required to adjust the rod to an absolute vertical position by manipulating a switch mounted in a handheld control box. The *S*s were first given two learning trials, after which they were given eight trials,³ using the standard in which the rod and frame were preset at standard positions used by Witkin et al. (1954). The score computed was the average number of degrees of deviation from the vertical for the eight trial series (Rod and Frame mean).

METHOD OF ANALYSIS

Of the 14 dependent variables that were derived from the perceptual procedures, no more than two scores were derived from any single test; there were no linearly interdependent subsets. Two methods were used to analyze these variables: (a) factor analysis was employed to determine whether the postulated perceptual dimensions, stimulus intensity, scanning, and field articulation, would emerge as

empirical factors, and (b) cluster analysis was employed to group *S*s on the basis of similarity of perceptual performance in order to determine whether the emergent groups would have the predicted characteristics.

The factor analysis was of the principal component type, with unities placed on the diagonal of the correlation matrix. Principal components were computed for each eigenvalue greater than .8, and all factors with eigenvalues of more than 1.0 were rotated to simple structure through the Varimax procedure. Individuals were assigned factor scores on each factor using the procedure described by Ryder (1965) and correlations computed between the factor scores and the independent variables.

RESULTS

Factor Analysis

Five factors emerged. The first we have interpreted as perceptual inefficiency. The perceptual tests with reliable loadings on this factor (greater than .40) include the rod-and-frame test; the two-flash threshold procedure; two measures of high anchoring effect on tasks involving discontinuous stimulus presentation—the Titchner Circles Illusion (D-A) and two-flash discrimination (D-A)—and the Petrie large block stimulation score. The *S*s with high scores on this factor evidenced greater error in rod-and-frame performance, higher two-flash thresholds, greater anchoring, and greater errors on the Petrie procedure. On the nonperceptual measures, they showed poor performance on the WAIS subtests and were rated high on the disorientation scale of the IMPS and poor in overall adjustment as measured by Montrose Rating Scales. The *S*s with high Factor 1 scores were older, were institutionalized longer, were more likely to have fathers in low-status occupations, and were more poorly educated. Thus this first factor appears to reflect a general dimension of efficiency-inefficiency, involving not only perceptual but also intellectual and social competence (Schooler & Parkel, 1966).

The second factor that emerged was interpreted as stimulus augmentation. It had loadings above .40 on both of the kinesthetic figural aftereffect measures (A) (B), on the visual size-estimation measure, and on the category width measure (see Table 1). Individuals scoring high on Factor II tended to enlarge (or at least not to attenuate) the experienced intensity of sensory stimuli and to

³ The rod-and-frame settings for the eight trials were: (1) Frame 242°, Rod 242°; (2) Frame 242°, Rod 298°; (3) Frame 298°, Rod 298°; (4) Frame 298°, Rod 242°; (5) Frame 242°, Rod 242°; (6) Frame 242°, Rod 298°; (7) Frame 298°, Rod 298°; (8) Frame 298°, Rod 242°.

TABLE 1
PERCEPTUAL FACTORS: LOADINGS

Measure	High score represents	I Perceptual inefficiency	II Stimulus augmentation	III Scanning	IV Anchoring	V Stimulus dependency
Visual size-estimation mean	Overestimation	.022	.438	-.629	.110	.069
Visual size-estimation D-A	Anchoring	-.029	-.315	-.660	-.146	-.141
Titchner mean	Minimal illusion	-.162	-.076	.650	.058	.257
Titchner D-A	Anchoring	.431	-.057	.510	.207	-.184
Müller-Lyer mean	Error	-.072	-.106	.117	-.224	.716
Müller-Lyer D-A	Anchoring	-.016	-.001	.087	.830	-.042
Rod-and-frame mean	Error	.594	-.021	.227	-.258	.497
Petrie large block stimulation	Overestimation	.677	-.022	.053	-.315	-.326
Petrie small block stimulation	Overestimation	-.127	.065	-.029	-.147	-.674
Two-flash mean	High threshold	.780	.014	-.130	.224	.279
Two-flash D-A	Anchoring	.612	-.232	-.322	.468	.080
Kinesthetic figural aftereffect ascending	Overestimation	.077	.882	.080	.017	-.119
Kinesthetic figural aftereffect total	Overestimation	-.228	.759	-.169	.042	-.085
Category width II	Broad categories	.066	.417	.231	.565	-.064
% of variance accounted		14.944	13.561	12.845	11.410	11.108

TABLE 2
PERCEPTUAL FACTORS: SIGNIFICANT CORRELATES

Measure	High score represents	I Perceptual inefficiency	II Stimulus augmentation	III Scanning	IV Anchoring	V Stimulus dependency
WAIS Digit Span	Better	-.433				-.366
WAIS Object Assembly	Correct	-.365	-.253			-.277
WAIS Similarities	Correct	-.577				-.242
WAIS Vocabulary	Correct	-.484				-.337
Chronicity	More	.383				.334
IMPS Apathy and Retardation	More			.233		
IMPS Paranoid Projection	More			-.281		
IMPS Perceptual Distortion	More			-.242		
IMPS Hostile Belligerence	More				.333	.251
IMPS Disorientation	More	.314		.332		
IMPS Conceptual Disorganization	More			-.293		
Staff Social	More			-.318		
Peer Social	More			-.308		
Wing Poverty of Speech	More		.246			.240
Wing Incoherent Speech	More			-.383		
Wing Flat Affect	More		.262			.345
Montrose Rating Scale	Good adjustment	.417				
Age (first admission)	Older		-.262		-.262	
Age (year born)	Younger		.300			
Education	Higher	-.304				-.301
Father's occupational status	Lower	.367				
Total time hospitalized	Longer	.441				.237

use broader categories in classifying perceptual stimuli. The Ss scoring high on this factor also performed poorly on the WAIS Object Assembly subtest. Apparently, classifying perceptual input in broad categories is associated with difficulty in correctly organizing the pieces of the Object Assembly task. In the interview situation, Ss with high Factor II scores were rated as evidencing flat affect and poverty of speech (Wing Scale). They were also younger at first admission and older at the time of testing.⁴

Factor III was interpreted as a scanning factor. Overestimation on the visual size-estimation task, which is significantly correlated with minimal scanning (Silverman & Gardner, 1967), has a very high negative loading on this factor. The Ss with high positive scores on this factor were regarded as relatively extensive scanners. They tended to underestimate and to anchor less on the visual size-estimation task, but anchored more on the Titchner Circles Illusion, despite being less prone to it (Table 1).

The Ss with high Factor III scores were rated on the IMPS as more likely to show apathy, retardation, and disorientation, and less likely to show conceptual disorganization, perceptual distortion (hallucinations), and paranoid projection. They were less likely to evidence incoherent speech (Wing Scale) and were also rated by ward attendants as socializing less with both staff and peers (Table 2). Apparently, individuals who extensively scan their surroundings are guarded both in their verbal communications and in the extent to which they interact with others.

Factor IV with its high loadings on Müller-Lyer (D-A), two-flash (D-A), and category width estimation was interpreted as an anchoring factor (Table 1). Individuals scoring high on this factor apparently employed broad categories in making perceptual judgments along specified dimensions and thus more readily perceived as identical phenomena that were similar but not identical. The Ss scoring high on this factor were younger on first ad-

mission and were rated on the IMPS as hostile (Table 2) and belligerent.

Factor V was interpreted as a "stimulus dependence" factor. The Ss scoring high on this factor evidenced high error scores on the rod-and-frame test and the Müller-Lyer Illusion. They also evidenced marked kinesthetic size underestimations on the Petrie procedure following stimulation with a small block. The loading of $-.326$ on the Petrie large block stimulation condition suggests that high-scoring Ss may also have a tendency toward stimulus-intensity reduction (Table 1). The negative, though relatively insignificant, loading of figural aftereffects (A) on this factor is consistent with the latter possibility. High-stimulus-dependent Ss evidenced poor performance on the WAIS subtests, although the correlations of the subtests with Factor V were somewhat lower than with Factor I. These Ss also were rated as hostile and belligerent (IMPS) and as evidencing flat affect and poverty of speech (Wing). Their educational level was lower and they had greater total hospitalization time (Table 2).

In summary, the factor analysis of the battery of perceptual tests indicates the existence of (a) the scanning, stimulus augmentation, and stimulus dependence factors specifically postulated in the tridimensional model (Silverman, 1964b, 1964c, 1967c); (b) an anchoring factor⁵ congruent with the findings of Boardman, Goldstone, Reiner, and Fathauer (1962); and (c) a perceptual inefficiency factor congruent with the findings of Johannsen, Friedman, and Liccione (1964) and Schooler and Scarr (1962).

Cluster Analysis

Cluster analyses of Ss using the hierarchical grouping method (Ward, 1963) were carried out on the performance scores for the 14 (non-redundant) perceptual variables. As noted previously, this form of analysis was carried out in order to examine the characteristics of the groups that emerge when Ss are classified

⁴ The apathy ratings and diagnosis correlates of Factor II did not quite reach statistical significance, although high Factor II scoring Ss tended to be rated as more apathetic and nonparanoid ($p < .06$).

⁵ This factor is also congruent with the works of M. Buchsbaum, J. Silverman, and R. Henkin, "Contrast Effects on the Auditory Evoked Response and Its Relation to Psychophysical Judgment," paper submitted for publication.

TABLE 3

HIERARCHICAL CLUSTERING OF SUBJECTS BASED ON PERCEPTUAL VARIABLES:
SIGNIFICANT DIFFERENCES BETWEEN CLUSTERS CONTAINING
THREE OR MORE SUBJECTS

Variable	7-cluster analysis				4-cluster analysis	
	1 ^a	2 ^b	5 ^c	6 ^d	1' ^e	2' ^f
Perceptual						
Size estimation (Mean)	5.16	5.26	5.83	4.85	5.04	5.34
Size estimation (D-A)	1.63	1.01	5.25	.31		
Titchner (D-A)	-.33	-.50	0.0	.75		
Müller-Lyer (Mean)	2.74	1.81	2.39	1.84	2.41	1.89
Müller-Lyer (D-A)	-6.09	-2.54	-2.23	-3.21		
Rod-and-frame (Mean)	10.53	7.62	17.08	26.27	15.49	8.85
Petrie large block stimulation	-.59	-1.40	1.77	.16		
Two-flash threshold (Mean)	53.66	83.25	118.33	103.75		
Two-flash threshold (D-A)	-18.67	5.00	50.00	31.25		
Kinesthetic figural aftereffects (A,A)	1.42	2.36	.50	-.69	.62	2.06
Kinesthetic figural aftereffects (Augmenting)	1.00	2.19	1.17	.08	.66	2.12
Category width II	29.20	38.90	28.67	28.75	28.38	37.57
Nonperceptual						
IMPS—Perceptual Distortion	3.13	4.50	24.00	5.00		
IMPS—Conceptual Disorganization	1.60	3.20	12.33	1.38		
WAIS—Similarities	12.87	14.80	7.67	7.50		
Wing—Incoherent Speech	1.27	1.60	3.67	1.25	1.25	1.87
Montrose Rating Scale	22.87	21.50	14.33	18.00		
Spohn Peer Social					2.60	5.00
Age (year born)	22.27	28.30	12.33	14.25	19.29	26.22
Age (first admission)					30.46	26.48
Factor I: Perceptual inefficiency	-.08	-.05	.20	.15		
Factor II: Stimulus augmentation	-.02	.06	-.05	-.13	-.07	.05
Factor III: Scanning	.01	-.03	-.31	.09	.04	-.07
Factor IV: Anchoring	-.11	.07	.01	.00	-.07	.06
Factor V: Stimulus dependency	.02	-.04	-.03	-.01		

^a Cluster 1, $N = 15$.

^b Cluster 2, $N = 20$.

^c Cluster 5, $N = 3$.

^d Cluster 6, $N = 8$.

^e Cluster 1', $N = 24$.

^f Cluster 2', $N = 23$.

according to their perceptual performances. The highest degree of within-group similarity and between-group difference occurred when the sample was broken down into seven clusters. Four of the clusters contained more than two individuals. These four form a very interesting pattern in terms of the differences among them in their factor scores. It can be seen from Table 3 that the clusters can be divided into two classes in terms of their perceptual efficiency scores, Clusters 1 and 2 being significantly more perceptually efficient than Clusters 5 and 6. Since Cluster 2 resembles Cluster 5 in that Ss in these clusters evidenced relatively minimal scanning, it is not surprising that when hierarchical clustering is continued from seven to four, two new clusters emerge containing more than two individuals. These two new clusters are the

result of collapsing the original four major clusters on the basis of similarities on the scanning dimension (Factor III) and by disregarding differences on the perceptual efficiency dimension (Factor I). Cluster 1' represents the combination of Clusters 1 and 6 with the addition of the one individual who comprised Cluster 4 ($n = 24$). Cluster 2' represents the combination of Clusters 2 and 5 ($n = 23$).⁶ Analysis of the differences between Ss in Clusters 1' and 2' indicates that Cluster 1' Ss evidenced significantly greater scanning (Factor III), narrower perceptual categories (Factor II), and less anchoring (Factor IV) (Table 3). On the individual perceptual tests, Cluster 1' Ss showed more error on the Müller-Lyer and rod-and-frame tests, less augmentation on the kinesthetic figural after-

⁶ See Footnote 4.

effects procedure (B), and narrower category widths on the Pettigrew procedure. They were also rated as evidencing less social interaction and less incoherent speech and were older both at time of testing and at first admission (Table 3). The differences in the perceptual performance of the individuals in these two clusters parallel the differences between the perceptual styles of paranoid and nonparanoid types described by Silverman (1964b, 1967c). The perceptual style seen as distinguishing the paranoid schizophrenics is similar to that manifested by Cluster 1' Ss. Comparison of the chart diagnoses of individuals in the two clusters reveals a greater, though not quite significantly so, proportion of paranoids in Cluster 1' (14 paranoids, 10 nonparanoids in Cluster 1', 11 paranoids and 12 nonparanoids in Cluster 2'). The picture is complicated, however, by the fact that the two clusters do not differ on any of the symptom variables that are conventionally used to distinguish paranoids from other schizophrenics (ratings of coherent delusions on the Wing Scale, and of paranoid projection, hostile belligerence, and grandiose expansiveness on the IMPS). The only significant differences in symptomatology between the clusters are the smaller amounts of incoherent speech and social interaction with peers shown by Cluster 1' Ss.

DISCUSSION

Before elaborating on the main findings of this investigation, some comments are in order regarding the different factor loadings found among a group of ostensibly similar variables—the Petrie large block and small block stimulation conditions and the kinesthetic figural aftereffect procedure employed by Silverman (1964a). On Factor I, errors of overestimation in the large block stimulation condition are associated significantly with other kinds of perceptual errors and with intellectual inefficiency. Performance scores on the small block stimulation condition do not load significantly on Factor I (Table 1). They do load, however, on Factor V where errors of underestimation in the small block stimulation condition are associated significantly with marked stimulus dependence. Note that on both factors, Ss evidencing a high degree of

error on the Petrie test are greatly influenced by the size of the interpolated stimulation bar employed. The fact that on Factor V underestimation errors on the large block stimulation condition also are associated with stimulus dependence (although to a lesser degree) suggests that consideration of the Petrie formulation of reduction-augmentation is necessary for an adequate conceptualization of the stimulus dependence factor. Indeed, this relationship between augmenting-reducing behavior and stimulus dependence has been specifically hypothesized elsewhere (Silverman, 1967a). Individuals who evidence marked reduction of the experienced intensity of stimulation as inferred from the Petrie procedure are expected to evidence a disposition to be less analytic or differentiated in their perceptual judgment, that is, to be more stimulus dependent. The fact that the kinesthetic figural aftereffect procedure employed by Silverman has its major loading on a different factor than the Petrie procedure is not comprehensible at this time. It may be that reduction-augmentation, inferred from figural aftereffect procedures, is affected by two different processes, one being related to analytic-differentiated responsiveness (Factor V), the other having to do with breadth of categorization behavior (Factor II). Such speculations are testable with a variety of techniques (Buchsbaum & Silverman, 1968), and an investigation of this specific problem is now being carried out in our laboratory.

In the main, the results of this investigation provide support for the existence of the three dimensions postulated in the original model (Silverman, 1967c). Stimulus intensity control, scanning control, and field articulation appear here in essentially the same form as stimulus intensity augmentation (Factor II), scanning (Factor III), and stimulus dependence (Factor V). In addition, two other factors emerged that we termed perceptual efficiency (Factor I) and anchoring (Factor IV). The perceptual efficiency factor, which accounted for the highest proportion of variance in this schizophrenic population, has not been found in comparable factor-analytic studies with normal Ss (Gardner, 1961; Gardner, Holzman, Klein, Linton, & Spence, 1959). It

emerged as the largest factor in this study quite probably because it taps the disruption of perceptual functioning, which is a prominent characteristic of many schizophrenic Ss. The emergence of anchoring, the second dimension not found in previous factor-analytic studies, is due quite likely to the inclusion of specific anchoring scores (D-A), which also were not included in previous relevant studies. The importance of this factor as a basic dimension of perceptual behavior is suggested by the high loading on the factor (.565) of the Pettigrew category width test. This positive correlation between anchoring scores and broad category width scores suggests the existence of a response dimension on which people differ in the extent to which they perceive stimuli as being equivalent, that is, as belonging to the same category.

Further support for the empirical foundation of the original tridimensional model is derived from the results of the cluster analysis. The pattern of differences between perceptual styles of Cluster 1' Ss and Cluster 2' Ss parallels the pattern of previously reported differences between paranoid schizophrenics and nonparanoid schizophrenics. As noted, however, the two clusters did not differ in prevalence of delusional symptoms that are conventionally accepted as the distinguishing characteristic of paranoid schizophrenia. Rather, the clusters differed in a group of symptoms that can be interpreted as interpersonal guardedness: Cluster 1' Ss (paranoidlike) exhibiting both more coherent speech in the clinical interview and less social interactions with staff and fellow patients. Thus, it could be that previously found relationships

between diagnosis of paranoid schizophrenia and the perceptual style evidenced by this reaction type were the result of a tendency to diagnose patients showing relatively guarded behavior and coherent speech as paranoid. The existence of such a relationship is supported by examination of some of the correlates of the scanning factor (Factor III) among the entire sample (Table 2). The Ss who scored high on this factor showed less incoherent speech and less social interaction with peers and staff. They also showed less conceptual disorganization, less perceptual distortion, and, in fact, significantly less paranoid projection on the interview-based IMPS. Whether this latter finding means that high scanners are actually less prone to paranoid delusions than low scanners or whether it merely reflects the fact that they are more guarded and less likely to expose their delusions in an interview situation remains an unsolved question. However, it is apparent that a highly significant relationship exists between extensive visual scanning of the environment and the group of traits that we have interpreted as behavioral guardedness.

Another noteworthy finding is the positive correlation obtained between Factor II, the stimulus augmentation factor, and flat affect on the Wing Scale (see Table 2). This correlation was, at first glance, somewhat confusing, for it was expected, on the basis of previous research (Silverman, 1967b), that schizophrenic Ss who augment the experienced intensity of stimulation would not tend to evidence flat affective responsiveness. The earlier work, however, dealt primarily with acute patients. Further analyses, in which acute and chronic patients were considered separately, yielded fascinating results. Acute patients with high stimulus augmentation scores showed significantly more anxious intropunitiveness and incoherent speech and significantly less disorientation (Table 4). In contrast, chronic patients with high stimulus augmentation scores showed significantly less anxious intropunitiveness and significantly more flat affect, poverty of speech, and disorientation. Thus, among acute schizophrenic patients, stimulus augmentation is related to the presence of inner turmoil, incoherence of

TABLE 4

CORRELATES OF STIMULUS AUGMENTATION (FACTOR II)
AMONG CHRONIC AND ACUTE PATIENTS

Scale	Acute ^a	Chronic ^b
IMPS Anxious Intropunitiveness	.538	-.340 ^c
IMPS Incoherent Speech	.623	.278
IMPS Disorientation	-.554	.337 ^c
Wing Flat Affect	.051	.360
Wing Impoverished Speech	.055	.394

^a $n = 15$. A correlation of .514 is needed for $p < .05$, two-tailed.

^b $n = 35$. A correlation of .335 is needed for $p < .05$, two-tailed.

^c The difference between the correlation for the two groups is $p < .005$, two-tailed.

speech, and minimal environmental disorientation. On the other hand, chronic schizophrenic Ss showing stimulus augmentation appear to have inner lives that are flat and undifferentiated. The distinct salience of external stimuli against such a flat background may make such stimuli seem stronger and differences between them irrelevant. In the two populations the same perceptual characteristic—stimulus augmentation—is thus associated with quite different affective response characteristics.

In summary, once certain fairly major qualifications are made, the results of the present study can be seen as supporting the original tridimensional model of perceptual functioning as it applies to hospitalized schizophrenics. One major qualification is an increase in the number of underlying dimensions; perceptual inefficiency and anchoring emerge as meaningful factors in addition to the three previously delineated ones—stimulus augmentation (stimulus intensity control), scanning (scanning control), and stimulus dependence (field articulation). Other modifications of the earlier theory made necessary by our findings are in the description of the interrelationships among certain of the personality and perceptual style variables. Behavioral guardedness replaces coherent paranoid delusions as a significant correlate of scanning. Stimulus augmentation and broad category width estimations are positively related to a high degree of anxious intropunitiveness only among acute patients. Among chronic patients the relationship is in the opposite direction.

Although these qualifications are substantial, and further research will probably result in even further modifications, the multidimensional description of perceptual functioning presented here has partially withstood the test of rigorous replication and certainly provides as promising a way of meaningfully accounting for the variance among schizophrenics as is now available.

REFERENCES

- BOARDMAN, W. K., GOLDSTONE, S., REINER, M. L., & FATHAUER, W. F. Anchor effects, spatial judgments, and schizophrenia. *Journal of Abnormal and Social Psychology*, 1962, 65, 273-276.
- BUCHSBAUM, M., & SILVERMAN, J. Stimulus intensity control and the cortical evoked response. *Psychosomatic Medicine*, 1968, 30, 12-22.
- GARDNER, R. W. Cognitive controls of attention deployment as determinants of visual illusions. *Journal of Abnormal and Social Psychology*, 1961, 62, 120-127.
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1(4).
- GARDNER, R. W., & LONG, R. I. Control, defence and centration effect: A study of scanning behavior. *British Journal of Psychology*, 1962, 53, 381-388.
- JOHANNSEN, W. J., FRIEDMAN, S. H., & LICCIONE, J. V. Visual perception as a function of chronicity in schizophrenia. *British Journal of Psychiatry*, 1964, 110, 561-570.
- LORR, M., & KLETT, C. J. Constancy of psychotic syndromes in men and women. *Journal of Consulting Psychology*, 1965, 29, 309-313.
- PETRIE, A. *Individuality in pain and suffering*. Chicago: University of Chicago Press, 1967.
- PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *Journal of Personality*, 1958, 26, 532-544.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-535.
- RACKOW, L. L., NAPOLI, P. J., KLEBANOFF, S. O., & SCHELLINGER, A. A. A group method for the rapid screening of chronic psychiatric patients. *American Journal of Psychiatry*, 1953, 109, 561-566.
- RYDER, R. G. Scoring orthogonally rotated factors. *Psychological Reports*, 1965, 16, 701-704.
- SCHOOLER, C., & PARKER, D. The overt behavior of chronic schizophrenics and its relationship to their internal state and personal history. *Psychiatry*, 1966, 29, 67-77.
- SCHOOLER, C., & SCARR, S. Affiliation among chronic schizophrenics: Relation to intrapersonal and birth order factors. *Journal of Personality*, 1962, 30, 178-192.
- SILVERMAN, J. Perceptual control of stimulus intensity in paranoid and nonparanoid schizophrenia. *Journal of Nervous and Mental Disease*, 1964, 139, 545-549. (a)
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 352-379. (b)
- SILVERMAN, J. The scanning control mechanism and "cognitive filtering" in paranoid and nonparanoid schizophrenia. *Journal of Consulting Psychology*, 1964, 28, 385-393. (c)
- SILVERMAN, J. Personality trait and "perceptual style" studies of the psychotherapists of schizophrenic patients. *Journal of Nervous and Mental Disease*, 1967, 145, 5-17. (a)
- SILVERMAN, J. Shamans and acute schizophrenia. *American Anthropologist*, 1967, 69, 21-31. (b)
- SILVERMAN, J. Variations in cognitive control and psychophysiological defense in the schizophrenias. *Psychosomatic Medicine*, 1967, 29, 225-251. (c)

- SILVERMAN, J. A paradigm for the study of altered states of consciousness. *British Journal of Psychiatry*, 1968, 114, 1201-1218.
- SILVERMAN, J., & GAARDER, K. Rates of saccadic eye movement and size judgments of normals and schizophrenics. *Perceptual and Motor Skills*, 1967, 25, 661-667.
- SUGARMAN, A. A., & CANCRO, R. Field independence and outcome in schizophrenia: A U-shaped relationship. *Perceptual and Motor Skills*, 1968, 27, 1007-1013.
- VENABLES, P. H. The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *Journal of Psychiatric Research*, 1963, 1, 279-287.
- WAPNER, S., & WERNER, H. *Perceptual development: An investigation within the framework of sensory-tonic theory*. Worcester, Mass.: Clark University Press, 1957.
- WARD, J. H. Hierarchical grouping to optimize an objective function. *Journal of American Statistics Association*, 1963, 58, 236-244.
- WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.
- WING, J. K. A simple and reliable subclassification of chronic schizophrenia. *Journal of Mental Science*, 1961, 107, 862-875.
- WITKIN, H. A. Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology*, 1965, 70, 317-336.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received May 1, 1968)

RELATIONSHIP BETWEEN PSYCHOSOCIAL ADJUSTMENT AND PERCEPTION OF MATERNAL ATTITUDES¹

LENORE BEHAR² AND ROGER F. SPENCER

University of North Carolina

A sample of 26 adult male hemophiliacs was divided into two groups on the basis of their psychosocial adjustment. The poorly adjusted group rated their mothers as significantly higher in tendencies toward dominating, overprotecting, and ignoring their children.

In studying the psychosocial adjustment of hemophiliacs, it was observed that these Ss frequently were exposed to similar patterns of child rearing. Support was found in earlier studies (Agle, 1964; Browne, Mally, & Kane, 1960) that mothers of hemophiliacs tend to overprotect and infantilize their sons. In fact, studies of juvenile diabetics (Swift, Seidman, & Stein, 1967) and of other children considered to be physically vulnerable (Green & Solnit, 1964) also revealed the mothers' tendencies to protect, infantilize, and dominate their children. These attitudes seemed initially based on the realistic possibility of physical danger to the child; however, even when the children in these studies were old enough to fend for themselves, the mothers continued their protective roles.

The mothers' attitudes in these studies seemed to reflect feelings of inadequacy as a childbearer and guilty feelings of having burdened a child with a chronic illness. Although feelings of guilt and responsibility were also apparent in mothers of diabetics and other chronically ill children, they may be more salient in mothers of hemophiliacs, for in this disease the defects in blood clotting are transmitted as sex-linked, recessive genetic traits by the carrier mother to the son. That is, the mother passes on a genetically determined condition which she herself does not have.

Hemophilia is a life-long, serious condition, found almost exclusively in males and usually

discovered in infancy or early childhood. It is characterized by the persistence of bleeding following an injury or by "spontaneous" bleeding into joints and soft tissues (Ratnoff, 1963). Although modern medical methods of handling bleeding episodes have decreased the mortality considerably, serious problems still occur—most frequently crippling of joints. In order to avoid the danger of bleeding, it may be necessary for a mother to restrict her son's physical activity, and in many situations to inhibit aggressive behavior. Having experienced these restrictions of childhood activity, coupled with frequent periods of being confined to bed and waited on, many adult hemophiliacs tend to have strong passive dependent tendencies (Agle, 1964; Browne, Mally, & Kane, 1960). Mattsson and Gross (1966) have found that the levels of adjustment achieved are closely related to child-rearing attitudes and anxiety in the mother.

Further investigation of the mother-child relationship should be fruitful in understanding the level of social adjustment achieved by adult hemophiliacs. Dealing with the adult patient offers the advantage of more crystallized personality characteristics. Understanding how he perceives his mother now and how he perceives her child-rearing techniques in retrospect should yield clues pertaining to social adjustment.

METHOD

In the present study, 26 male Ss were divided into two groups on the basis of their psychosocial adjustment, which was measured by the Social Maturity Index of the California Personality Inventory (Gough, 1966). The scores were divided at the median, yielding 14 Ss rated as having low social maturity and 12 as having high social maturity. In this sample the range of scores was 40-58, with a median of 47. The mean of the low group was 45.57

¹The study was supported by Medical School Grant 324 UNI-1-218 from the University of North Carolina.

²Requests for reprints should be sent to Lenore Behar, Department of Psychiatry, University of North Carolina Medical School, Chapel Hill, North Carolina 27514.

TABLE 1
DESCRIPTION OF THE SUBJECT GROUPS

Item	Group	
	High social maturity (<i>n</i> = 12)	Low social maturity (<i>n</i> = 14)
Age	<i>M</i> = 29.0 (range 16-75)	<i>M</i> = 32.9 (range 17-67)
Social class*	<i>M</i> = 3.5	<i>M</i> = 4.0
Education	<i>M</i> = 12th grade	<i>M</i> = 10th grade
Severity of illness	4 mild 8 severe	5 mild 9 severe

* Measured on a 5-point scale, with 1 representing the highest class (Hollingshead & Redlich, 1958).

and that of the high group, 53.16, the difference being statistically significant beyond the .001 level ($t = 8.56$). In other respects, the two groups were rather equally matched, as seen in Table 1.

The Ss in both groups were administered the Mark Scale (Harris, 1955), a scale designed to measure S's perception of his mother's child-rearing attitudes. The scale consists of 52 items, items which were rated by a panel of experts to reflect dominating, overprotecting, and ignoring on the part of the person being rated—in this case the mother and the ideal mother. On this test, S is asked whether his mother would agree or disagree with the following types of statements:

A mother should never be separated from her child.

If a child is quiet for a few minutes, a mother should immediately find out what he's thinking.

RESULTS AND DISCUSSION

Table 2 shows the differences between the two groups on this questionnaire. The high-social-maturity group consistently rate the

TABLE 2
PERCEPTION OF MATERNAL ATTITUDES: MEAN NUMBERS OF ITEMS INDICATING NEGATIVE CHILD-REARING PRACTICES

Rating of	Group		<i>t</i>
	High social maturity (<i>n</i> = 12)	Low social maturity (<i>n</i> = 14)	
Real mother			
Dominance (15 items)	7.08	9.64	3.01*
Ignoring (13 items)	3.83	6.36	3.18*
Overprotecting (24 items)	5.91	9.26	2.66*
Ideal mother			
Dominance (15 items)	5.58	9.71	3.97**
Ignoring (13 items)	3.67	6.07	2.97*
Overprotecting (24 items)	5.83	10.86	2.57*

* $p < .05$.

** $p < .01$.

mother as less dominating, ignoring, and overprotecting than the low-social-maturity Ss did. The more mature Ss rated their mothers as better mothers, or at least as mothers lower in these three negative characteristics. Interesting, too, is the direction of change in rating the ideal mother, who was described to Ss in the instructions as "the best mother possible." The more socially mature Ss rated the ideal mother even lower than the real mother. The less socially mature Ss rate the ideal mother even higher on these negative characteristics, especially on the dimension of overprotection. One might infer the statement, "My mother was quite overprotecting but an ideal mother should be even more so" by the less mature S. We feel that the ideal mother represents fantasied wish fulfillment and that these individuals would have wanted even more overprotection and domination than their mothers provided. Support for this interpretation is gained from previous work with this same S group by Spencer (1968), in which he found that many of the less well-adjusted patients have had serious marital problems or have been unable to cut dependent ties with their mothers, some have developed hostile-dependent relationships with their wives, and some have had continued problems with alcoholism—all reflecting in one way or another a passive dependent orientation with a wish to be dominated and protected.

The present study does not indicate whether a person is less well adjusted because he has been raised by a dominating and overprotective mother or whether he just sees his relationship with his mother in this way because he is poorly adjusted. However, the previously mentioned studies of hemophiliacs lend considerable support to the former interpretation, that a dominating and overprotective mother tends to produce the type of adult described above—passive and dependent. If so, preventive work is necessary. Certainly, the mother-child relationship in the case of genetically determined conditions or chronic childhood disorders has more of a chance of being problematic than when the child is normal. At least in these trouble-prone cases, early counseling with the mothers or

with both parents could be offered in an attempt to alleviate feelings of guilt and to thwart the development of an overprotective relationship.

REFERENCES

- AGLE, D. P. Psychiatric studies of patients with hemophilia and related states. *Archives of Internal Medicine*, 1964, 114, 76-82.
- BROWNE, J. W., MALLY, M. A., & KANE, R. P. Psychosocial aspects of hemophilic children and their families. *American Journal of Orthopsychiatry*, 1960, 30, 730-740.
- GREEN, M., & SOLNIT, A. J. Reactions to the threatened loss of a child: A vulnerable child syndrome. *Pediatrics*, 1964, 34, 58-66.
- GOUGH, H. G. Appraisal of social maturity by means of the CPI. *Journal of Abnormal Psychology*, 1966, 71, 189-195.
- HARRIS, J. G. A study of the mother-son relationship in schizophrenia. Unpublished doctoral dissertation, Duke University, 1955.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- MATTSSON, A., & GROSS, S. Social and behavioral studies on hemophilic children and their families. *Journal of Pediatrics*, 1966, 68, 952-964.
- RATNOFF, O. D. The therapy of hereditary disorders of blood coagulation. *Archives of Internal Medicine*, 1963, 112, 92-111.
- SPENCER, R. F. Incidence of social and psychiatric problems in a group of hemophiliac patients. *North Carolina Medical Journal*, 1968, 29, 332-335.
- SWIFT, C. R., SEIDMAN, F., & STEIN, H. Adjustment problems in juvenile diabetes. *Psychosomatic Medicine*, 1967, 29, 555-571.

(Received May 4, 1968)

COGNITIVE INTERFERENCE BY NONVERBAL SYMBOLS IN SCHIZOPHRENICS

JONAS LANGER,¹ KENNETH B. STEIN

University of California, Berkeley

B. G. ROSENBERG

Bowling Green State University

The degree of cognitive interference evidenced in schizophrenics, as compared with normal medical controls, on the Color-Phonetic Symbol Test (a non-conventional version of the Stroop Word-Color Interference Test) was assessed. As expected, the performance of schizophrenics showed significantly greater cognitive interference. Two reasons for these results were noted, one in terms of general personality concepts and the other in terms of a cognitive theory of symbolic processes.

Phonetic symbols may influence the performance of a different but related cognitive task in ways that parallel the effect of verbal symbols (Langer & Rosenberg, 1966). In a study modeled after the conventional Word-Color Interference Test or WCIT (e.g., Stroop, 1935), phonetic symbols were printed in different colors to create a congruent (C) and noncongruent (N) card. On Card C the phonetic symbols were always printed in the preferred choices, that is, the colors with which most Ss matched them, for example, *zah* in red ink and *tur* in green ink, as determined by Langer and Rosenberg (1964). On Card N the phonetic symbols were always printed in a nonpreferred color, for example, *zah* in green ink and *tur* in red ink. As in the conventional test, the task on the Color-Phonetic Symbol Test (CPST) was to name the ink colors as rapidly and accurately as possible. It was found that the noncongruent phonetic symbols interfered with the speed and accuracy of the performance on Card N as compared with Card C.

The present investigation further tested (a) the general assertion that the influence of nonconventional symbolic activity may parallel that of conventional symbolization upon cognitive activity, and (b) the particular assertion that nonconventional phonetic symbols may interfere with the maintenance of directed attention toward the performance of

another cognitive task (Langer & Rosenberg, 1966). It was expected that the performance of schizophrenics on the CPST would parallel their performance on the WCIT. Since schizophrenics evidence greater interference than normals on the WCIT (Wapner & Krus, 1960), it was hypothesized that they also evidence greater interference on the CPST.

METHOD

Subjects

Two groups of Ss were tested on the CPST, an experimental and a control group. The experimental group consisted of 20 recently hospitalized acute schizophrenic patients (10 males and 10 females). The control group consisted of 20 newly admitted medical patients (10 males and 10 females) hospitalized for nonpsychiatric reasons.

The groups were very closely matched for age (schizophrenics' mean age was 35 yr., $SD = 13.6$; medicals' mean was 38 yr., $SD = 13.6$) and education (schizophrenics' mean was 12.6 yr., $SD = 1.66$; medicals' mean was 12.3 yr., $SD = 1.19$). Neither the difference in age nor in education was statistically significant.

Procedure

All Ss were tested on the CPST, which is described fully in Langer and Rosenberg (1966). On Card C, 10 phonetic symbols are printed in the colors with which they were most consistently matched by a normative group, that is, red—*zah*, *klak*, *skaf*; blue—*mumle*, *oom*, *sool*, *mu*; green—*tur*, *ish*, *nerd*. On Card N the same symbols are printed in colors with which they were seldom matched. Each card consists of 10 rows of 10 symbols each for a total of 100. The measure used was the length of time to name the colors per card. In order to balance possible order effects, half of each group was

¹Requests for reprints should be sent to Jonas Langer, Department of Psychology, University of California, Berkeley, California 94720.

tested on Card C first and then Card N, while the other half of each group was tested in reverse order.

RESULTS

Both groups of Ss evidenced the expected interference effect on the CPST. In all, 39 of the 40 Ss tested took a longer time to complete Card N than C; no S took longer on Card C; and 1 medical S took an equal time on both cards. This finding serves as a further replication of previous findings (Langer & Rosenberg, 1966; Stein & Langer, 1966) that the interference effect of the CPST is a general phenomenon.

Table 1 presents the mean duration by groups to complete Cards N and C. As can be seen, the mean difference between duration to complete the two cards by the schizophrenic group was more than twice that of the medical control group.

Table 1 also reveals that the schizophrenics were slower than the medical controls on both cards. In order to adjust for this initial mean duration difference between groups, each S's congruent score was used as his base line so that his percentage change from C to N could be calculated as follows: $\text{noncongruent} - \text{congruent} / \text{congruent} \times 100$. The percentage mean difference between duration on Card N and Card C was then calculated for each group, and the percentages are presented in Table 2. Multiple *t* tests were performed. Only the predicted difference between schizophrenics and normals was statistically significant.

DISCUSSION

As expected, the noncongruity between the phonetic symbols and the colors clearly interfered with the color-naming performance to a significantly greater extent among schizo-

TABLE 1

OVERALL MEAN DURATIONS ON THE CPST

Group	<i>M</i> (sec.)		
	Noncongruent	Congruent	Difference
Schizophrenics	142.2	112.4	29.8
Medicals	91.1	77.5	13.6
Both	116.6	94.9	21.7

TABLE 2
PERCENTAGE MEAN DIFFERENCES

Item	<i>M</i>	<i>SD</i>	<i>t</i>
Schizophrenics	28.2	18.8	2.5*
Medicals	16.7	6.9	
Males	24.3	15.2	0.53
Females	20.6	15.1	
N to C	25.7	16.3	0.68
C to N	19.2	13.4	

Note.—Since time scores tend to be positively skewed, the percentage time differences were transformed into logarithmic scores. It was found that the probabilities for significance remained the same.

* $p < .025$, one-tailed test.

phrenics than among normals. Two different (but hardly opposing, and possibly complementary) reasons may be given to begin to explain this comparative phenomenon.

The first reason is that cognitive interference is directly related to intra- and inter-personal effectiveness. Previous studies on overt interference on the conventional WCIT (e.g., Wapner & Krus, 1960) and covert interference on the nonconventional CPST (Stein & Langer, 1966) have demonstrated that high interference is associated with poor personal adjustment. Consequently, it is to be expected that the low personal effectiveness of schizophrenics will be reflected in greater interference on the CPST, just as it is on the WCIT.

The second reason is more directly cognitive in nature. It has been previously demonstrated that (a) low consensual agreement in matching phonetic symbols to their color referents leads to increased interference on the CPST (Langer & Rosenberg, 1966), and (b) the consensual agreement of schizophrenics in matching phonetic symbols with their color referents is about half that of normals (Rosenberg & Langer, 1965). Low consensuality has been interpreted to reflect the person's relative lack of cognitive competence in differentiating and keeping separate symbols from their referents, and that should consequently lead to greater interference of performance (see Langer & Rosenberg, 1966, for a fuller exposition of the argument). Since schizophrenics are low consensualizers, it is to be expected, therefore, that they will evidence relatively more interference than normals on the CPST.

REFERENCES

- LANGER, J., & ROSENBERG, B. G. Nonverbal representation of verbal referents. *Perceptual and Motor Skills*, 1964, 19, 363-370.
- LANGER, J., & ROSENBERG, B. G. Symbolic meaning and color naming. *Journal of Personality and Social Psychology*, 1966, 4, 364-373.
- ROSENBERG, B. G., & LANGER, J. A study of nonverbal symbolic behavior of schizophrenics. *Journal of Clinical Psychology*, 1965, 21, 243-247.
- STEIN, K., & LANGER, J. The relation of covert cognitive interference in the Color-Phonetic Symbol Test to personality characteristics. *Journal of Personality*, 1966, 4, 241-251.
- STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, 18, 643-662.
- WAPNER, S., & KRUS, D. M. Effects of lysergic acid diethylamide, and differences between normals and schizophrenics in the Stroop Color-Word Test. *Journal of Neuropsychiatry*, 1960, 2, 76-81.

(Received May 15, 1968)

CORRELATES OF EMPATHY IN PSYCHOTHERAPY: A REPLICATION

ALLEN E. BERGIN¹ AND LAWRENCE G. JASPER

Teachers College, Columbia University

This study replicates previous findings regarding the relationship between various therapist personal characteristics and therapist empathy levels during psychotherapy. It was concluded that (a) there is a moderate inverse relationship between therapist MMPI disturbance levels and degree of therapist empathy in interviews; (b) there is no correlation between indexes of therapists' intellectual ability or academic achievement and their empathy scores; and (c) there is some reason to doubt that Truax Empathy scores are related to patient outcome in nonclient-centered therapy.

An elaborate series of studies has been published in recent years which gives support to Rogers' (1957) theory that therapist empathy is a correlate and cause of positive personality change. Some of this evidence also suggests that low empathy is related to client deterioration (Bergin, 1963, 1966; Strupp & Bergin, 1969; Truax & Carkhuff, 1967).

If these conclusions are correct, an obvious question to pose is: "What produces variation in therapist empathy? Is it therapist personal characteristics, therapist training and experience, or client characteristics?" Evidence to date indicates that all three factors are related to level of therapist empathy (Bergin & Solomon, 1968; Truax & Carkhuff, 1967), though no comparison of their relative contributions to empathy has been made.

The present study replicates and extends an earlier study in which therapists' personality test scores were found to be significantly correlated with empathy levels, as rated from tape recordings of treatment sessions (Bergin & Solomon, 1968). Both clinical and research evidence suggest that personal characteristics of therapists affect treatment outcomes and, particularly, that personality disturbance in the therapist interferes with therapeutic efficacy (for reviews see Bergin, 1967; Strupp & Bergin, 1969).

STUDY I

In Study I (Bergin & Solomon, 1968), 18 postinternship students in clinical and coun-

seling psychology and 26 of their clients were studied. The main data consisted of correlations between therapists' MMPI and Edwards Personal Preference Schedule (EPPS) scores and their empathy levels during therapy, as measured by the Truax Accurate Empathy Scale.

It was originally predicted that the MMPI *D*, *Pt*, *Hy*, and *Pd* scales would correlate negatively with empathy, and that EPPS Nurturance and Intracception would correlate positively. Study I (see Tables 2, 3, and 4) yielded significant negative correlations between therapists' empathy and their depression (*D*) and anxiety (*Pt*) scores on the MMPI. Scores from the EPPS Consistency and Intracception scales correlated negatively with empathy while Dominance and Change correlated positively. Chronological age and supervisor ratings of therapist competence also correlated positively, whereas grades and Graduate Record Examination (GRE) scores were unrelated to empathy.

STUDY II

Method

Study II replicated Study I exactly except for the addition of the several measures noted.

Clients and Therapists

The clients ($n = 48$) were late adolescent and adult neurotics who were seen in an intensive psychotherapy practicum during each of 2 academic years in the Teachers College Psychological Consultation Center.

The therapists ($n = 36$) were fourth-year graduate students in clinical and counseling psychology who had seen a few cases for diagnosis and brief therapy

¹ Requests for reprints should be sent to Allen E. Bergin, Department of Psychology, Teachers College, Columbia University, New York, New York 10027.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THERAPIST
VARIABLES: STUDY II

Variable	M	SD	Variable	M	SD
Empathy ratings	3.16	2.03	EPPS		
Supervisor rating	5.31	1.30	Con	53.11	9.95
Grade Average					
Practice	6.79	1.03	Aff	48.25	9.89
Academic	6.62	.52	Def	44.00	8.27
GRE					
Psychology	598.53	66.11	Ach	61.17	6.59
Quantitative	563.53	90.80	Ord	42.80	9.67
Verbal	653.24	75.55	Exh	49.36	9.10
Age					
23-50	29.08	6.19	Aut	54.81	9.56
23-29	25.48	1.62	Int	53.42	8.01
MMPI			Suc	53.53	10.10
K	55.70	6.74	Dom	47.17	10.40
Hs	48.42	6.63	Aba	40.77	9.15
D	57.17	9.11	Nur	49.17	8.55
H _y	57.92	6.41	Chg	49.17	10.00
P _d	59.17	9.60	End	46.08	10.74
M _h					
Male	33.00	4.16	Het	57.44	10.34
Female	43.05	2.58	Agg	51.44	9.33
P _d	58.42	6.44			
P _t	56.08	7.63			
S _c	56.86	8.86			
M _a	56.56	7.94			
S _i	51.61	7.87			
E _s	65.17	7.13			

* Raw scores.

for 1 previous year in the same clinic and who also had completed a 1-yr. full-time internship.

Personality. MMPI and EPPS scores were again used to assess level of disturbance and general personality characteristics. The therapists were not highly familiar with the scoring and use of these inventories. Special coding and other precautions were taken to insure anonymity of respondents, and checks on validity of profiles revealed no evidence of faking.

Empathy. Therapists regularly recorded their sessions and provided 1 or 2 hr. on each of two patients^a after the request for data was made. Most of the tapes thus collected were obtained from therapy during the week preceding or succeeding the announcement and, therefore, there was little chance for a biased selection of tapes being sampled.

Empathy was rated on the middle one-third of each hour sampled on the Bergin-Solomon revision of the Truax Accurate Empathy Scale (see Truax & Carkhuff, 1967, p. 59).

Other. Ratings were also obtained from supervisors of the general therapeutic competence of each student, as well as GRE scores, grade-point averages, and five therapist self-ratings on their personal intent to be empathic, interpretive, supportive, advising, and warm.

Client Measures

No measures of client behavior or characteristics were obtained except for a global outcome rating

² Several therapists were unable to provide tapes on two patients; therefore, their empathy ratings are based on one case only. This is accepted in light of the fact that therapist empathy scores across patients in the Bergin-Solomon study correlated .98.

on part of the sample made by the supervisor after discussion with the therapist and the three other students involved in that supervisory group. These ratings were made on a 7-point scale ranging from "considerably improved" to "considerably worse."

RESULTS

Interrater reliability on the empathy ratings was based on 32 therapy hours which were rated by two judges independently. The Pearson correlation between ratings was .88.

The prime object of the replication was to cross-validate the MMPI and EPPS findings from the Bergin-Solomon study. Table 1 includes means and standard deviations for all variables, Table 2 shows the correlations between empathy and the therapist personality scores which were the focus of the replication, and Table 3 presents similar data for all other personality measures.

Table 2 reveals a rather clear replication of the original MMPI data. The *D* and *P_t* scales correlate negatively and significantly with empathy. The EPPS correlations do not replicate in any respect, although several new correlations appear (Table 3). It may be safe to assume that in both instances the EPPS correlations were simply random fluctuations. It is worth noting that in both Study I and Study II, practically all MMPI pathology scales correlated negatively with empathy even though they were not significant.

Table 4 includes correlations of other variables with empathy.

TABLE 2

SIGNIFICANT PERSONALITY CORRELATES OF EMPATHY
FROM STUDY I AND REPLICATED CORRELATIONS
FROM STUDY II

Test	Study I ^r	Study II ^r
MMPI ^a		
<i>D</i>	-.41*	-.31**
<i>P_t</i>	-.54*****	-.38****
EPPS ^b		
Con	-.54***	.03
Int	-.53***	.01
Dom	.54***	.16
Chg	.55***	.14

Note.—One-tailed tests were used where the observed relationship was in the hypothesized direction.

^a *n* = 12 in Study I, *n* = 36 in Study II.

^b *n* = 13 in Study I, *n* = 36 in Study II.

* *p* < .06, one-tailed.

** *p* < .05, one-tailed.

*** *p* < .05, two-tailed.

**** *p* < .025, one-tailed.

***** *p* < .01, one-tailed.

TABLE 3

PERSONALITY MEASURES UNCORRELATED WITH EMPATHY IN STUDY I, AND STUDY II CORRELATIONS FOR THE SAME VARIABLES

Test	Study I r	Study II r
MMPI ^a		
<i>K</i>	.15	.04
<i>Hs</i>	-.31	-.24
<i>Hy</i>	-.30	-.20
<i>Pd</i>	.10	-.23
<i>Mf</i>	.37	
Male ($n=16$)		.13
Female ($n=20$)		.34
<i>Pa</i>	-.05	.03
<i>Sc</i>	-.05	-.11
<i>Ma</i>	.32	-.07
<i>Si</i>	-.18	-.11
<i>Es</i>	.08	.32
EPSP ^b		
Aff	-.18	-.38**
Def	-.23	-.22
Ach	-.15	.34*
Ord	-.41	.07
Exh	-.12	.06
Aut	.37	.01
Suc	.00	.02
Aba	-.12	-.37*
Nur	-.19	-.27
End	.08	.17
Het	-.05	.08
Agg	.04	-.02

^a $n = 12$ in Study I, $n = 36$ in Study II.

^b $n = 13$ in Study I, $n = 36$ in Study II.

* $p < .05$, two-tailed.

** $p < .02$, two-tailed.

Outcome Measures

For a limited number of therapists ($n = 20$) and patients ($n = 24$) measures of patient improvement were available, as described in the Method section, and were found related neither to empathy ($r = .05$) nor to therapist personality characteristics (D , $r = -.05$; Pt , $r = .08$).

DISCUSSION

Personality Characteristics

The most important finding is that therapists' Depression (D) and Anxiety (Pt) correlate negatively with their empathy in the live therapeutic process. This is clearly in keeping with the common viewpoint that therapist personal disturbances interfere with therapeutic effectiveness. While we cannot yet set up cutoff scores for selecting out or identifying therapists whose personal disturbances interfere with clinical performance, we have established that there is a moderate inverse

relationship between level of disturbance and level of competence in one important dimension of therapeutic behavior. This confirms the importance of focusing upon therapist characteristics in the study of treatment processes and outcomes. Provided that further studies substantiate the assumed relationship between empathy and outcome, such inquiry is likely to clarify the causes of the large variation in observed therapeutic effects, which includes both positive changes and significant deterioration among therapy patients.

The failure of EPSP scores to correlate with in-therapy empathy ratings may be due to a general limited validity of this instrument or to its lesser relevance to therapeutic phenomena. We cannot interpret the fluctuating correlations with this inventory as being other than due to chance.

Intellectual Characteristics

The lack of relationship between empathy and indexes of academic and intellectual competence are not surprising. They simply reaffirm the almost universally held view that scholastic abilities are not very relevant to effectiveness in establishing therapeutic relationships with clients. The nonsignificant results therefore have considerable psychological significance.

Intellectual capacities may, of course, be of value in developing expertise in other

TABLE 4
OTHER CORRELATES OF EMPATHY

Variable	Study I		Study II	
	r	N	r	N
Supervisor rating	.41*	18	-.40	18
Grade average				
Diagnostic practice	-.17	18	.01	23
Academic	-.16	18	.07	35
GRE				
Psychology	-.18	17	.33**	34
Quantitative	.21	18	-.02	34
Verbal	-.30	18	.11	34
Age				
23-50	.15	17	-.04	23
23-29	.50*	11	-.15	36
Hours of diagnostic and therapeutic experience	—	—	.28	19

* Not available.

* $p < .05$, one-tailed.

** $p < .05$, two-tailed.

forms of therapy, and their lack may be critical below a certain low level rarely found among graduate students. It may also be that the range of grades and intelligence for these individuals was restricted enough to reduce the possibility of obtaining significant correlations.

Outcome Data

The failure to find any correlation between empathy scores and outcome ratings runs contrary to the large number of studies published by the Truax group. At first, this might be interpreted as being due to the crudity of the present outcome ratings; however, the same ratings yielded significant findings in a coordinate study by Feitel (1968). Using the same sample of therapists and patients, she found that patient self-ratings of feeling understood by the therapist correlated .59 with outcome and that two patient ratings of therapist behavior from the Barrett-Lennard Relationship Inventory also related to outcome (Empathic Understanding, .34; Regard plus Unconditionality, .77). None of these ratings was correlated with Truax Empathy scores.

These findings raise at least tentative questions about the generalizability of Truax's findings to nonclient-centered therapy, and they suggest that his procedures should be replicated in full by independent investigators on new samples. Truax did attempt a replication in cooperation with the Jerome Frank group at Johns Hopkins Hospital in Baltimore (Truax, Wargo, Frank, Imber, Battle, Hoehn-Saric, Nash, & Stone, 1966). However, the results of that study actually were quite equivocal in spite of contrary assertions. We are presently investigating this issue in our own center.

Additional Issues

It is evident that in both the original sample and in the replication we have obtained a few significant results from a large number of correlations computed on small n 's. Ordinarily the meaning of such results would be ambiguous and their value small; however, there are good reasons for believing that the data have psychological significance: (a) Many of the correlations were computed

simply out of curiosity and were unessential to the main task, for example, of the 56 correlations with the MMPI and EPPS only 12 were technically required by the design, 6 to test the original hypotheses and 6 more to cross-validate the original correlations. It was thought useful to present a full correlation table, but it is worth noting that 2 of the original 6 relationships held up under replication. (b) Those findings which did hold make very good sense theoretically and they appear on the MMPI, which has consistently correlated with various aspects of the therapeutic process (Strupp & Bergin, 1969). (c) In addition, many of the zero correlations are of obvious importance. They are in keeping with the views of many professionals (e.g., no relationship between GRE and Empathy), and should not be considered as necessarily detracting from the significance of the results.

Several additional variables concerning age, experience level, therapist and supervisor orientation, and supervisor evaluation of the therapist were studied. While a number of suggestive leads occurred in these results, no unambiguous findings occurred and thus they are not reported. Possible causes of these results and of the only modest correlations occurring in the main data are (a) that the Truax Empathy scale is currently only a very crude way of approximating how therapists actually behave and that it may not adequately measure types of empathic communication which occur outside the client-centered framework, (b) that the therapist sample was relatively homogeneous, and (c) that some variables were thus restricted in range.

REFERENCES

- BERGIN, A. E. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, 1963, 10, 244-250.
- BERGIN, A. E. Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology*, 1966, 71, 235-246.
- BERGIN, A. E. An empirical analysis of therapeutic issues. In D. Arbuckle (Ed.), *Counseling and psychotherapy: An overview*. New York: McGraw-Hill, 1967.
- BERGIN, A. E., & SOLOMON, S. Personality and performance correlates of empathic understanding in psychotherapy. In T. Tomlinson & J. Hart (Eds.), *New directions in client-centered therapy*. Boston: Houghton-Mifflin, 1968.

- FEITEL, B. Feeling understood as a function of a variety of therapist activities. Unpublished doctoral dissertation, Teachers College, Columbia University, 1968.
- ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 1957, 21, 95-103.
- STRUPP, H. H., & BERGIN, A. E. Some empirical and conceptual bases for coordinated research in psychotherapy. *International Journal of Psychiatry*, 1969, 7, 18-90.
- TRUAX, C. B., & CARKHUFF, R. R. *Toward effective counseling and psychotherapy*. Chicago: Aldine, 1967.
- TRUAX, C. B., WARGO, D. G., FRANK, J. D., IMBER, S. D., BATTLE, C. C., HOEHN-SARIC, R., NASH, E. H., & STONE, A. R. Therapist empathy, genuineness, and warmth and patient therapeutic outcome. *Journal of Consulting Psychology*, 1966, 30, 395-401.

(Received May 27, 1968)

EFFECT OF DEVELOPMENTALLY RELEVANT INFORMATION AND SEVERITY OF DISTURBANCE ON THE UNDERSTANDING AND EVALUATION OF DEVIANT BEHAVIOR

MARVIN R. GOLDFRIED¹

State University of New York at Stony Brook

ALLEN A. ADINOLFI

University of Rochester

AND EARL DWORKIN²

Schenectady County Child Guidance Center, New York

To test the assumption that attitudes toward deviant behavior could be influenced by the degree of disturbance as well as understanding of the behavior, 30 Ss evaluated individuals manifesting various degrees of deviant behavior. Another group of 30 Ss evaluated the same individuals, but were also provided with developmentally relevant material to help them understand the deviant behavior. The results indicated that Ss in both groups tended to react negatively to individuals manifesting more deviant behavior patterns. Although the availability of developmentally relevant material increased the understanding of the deviant behavior, the negative evaluations remained unchanged. Some minimal evidence is provided for the possibility that an understanding of moderately deviant behavior discourages the use of the "mental illness" label.

In recent years workers in the mental health professions have shown a growing interest in the public's perception of "mental illness." Of particular concern is the consistent negative attitude permeating the public opinion of deviant behavior (Nunnally, 1961; Sarbin, 1967; Szasz, 1961), along with the view that mental illness represents a "disorder without a cause" (Cumming & Cumming, 1957; Starr, 1955). To the extent that deviant behavior is viewed as being strange, unmotivated, and incapable of being understood, it may only serve to add to the deviant individual's self-evaluation, as well as a feeling that because he has some sort of "mental illness," there is little that he can do to change his own behavior.

The generally assumed deleterious effect of the negative public attitude on the stability of the discharged mental patients' readjustment to society has led to the study of factors which might be relevant to the induction of a more favorable attitude toward deviant and

maladaptive behavior. For example, Rothaus, Hanson, Cleveland, and Johnson (1963) found that a "problems in living" set (cf. Szasz, 1961) in employment interviews led to a higher likelihood of employment for patients than did a mental illness set.

While studies such as those by Rothaus et al. demonstrate a practical consequence of a non-mental-illness set, a more prevalent assumption has been that an increase in information concerning deviant behavior could be effective in fostering a more favorable attitude toward the mental patient. A phenomenon often observed in clinical settings is the tendency to become more sympathetic and less negative in one's evaluation of deviant and maladaptive behavior once the causes of these behaviors are known. Similarly, Yarrow, Schwartz, Murphy, and Deasy (1955) have noted that the wife of a mental patient typically attempts to "normalize" her husband's deviant behavior by looking for reasons underlying such behavior, and Starr (1955) has further suggested that as long as behavior is made intelligible, there is a tendency to regard deviant behavior as something other than mental illness.

It seems, however, that while increased information may induce a change in public

¹ Requests for reprints should be sent to Marvin R. Goldfried, Department of Psychology, State University of New York, Stony Brook, Long Island, New York 11790.

² The authors would like to thank Barry Sherman for some of his comments during the early phases of this study.

reaction to the mental patient, the nature of this attitudinal change has not yet been determined. For example, Altrocchi and Eisdorfer (1961) have shown that although completion of a course in abnormal psychology leads to more understanding of the mental patient, it does not induce a more favorable attitude toward him. This lack of attitudinal change is attributed to the fact that Ss in their study were all college and nursing students, and consequently were fairly well-informed prior to the course.

In an attempt to improve attitudes toward people displaying deviant behavior, Nunnally (1961) varied the label that was used to introduce the description of these behaviors. Interestingly enough, he found that semantic differential ratings of a description of *the same person* would vary, depending on whether he was labeled as "an insane man," "an emotionally disturbed man," or simply "a man."

In contemporary discussions of the concept of mental illness, a growing dissatisfaction with the medical model has become most evident (Maher, 1966; Szasz, 1961; Ullmann & Krasner, 1965). The conceptualization of deviant behavior as the manifestation of an underlying disease, in addition to being questionable from a scientific, explanatory viewpoint, may have negative effects on the public's attitudes toward this deviant behavior by making it more incomprehensible. By way of contrast, we would expect that the psychological model of maladaptive behavior, which emphasizes the role of social-learning history in the explanation of concurrent behavior patterns—adaptive as well as maladaptive—could very well add to the public's understanding as well as a more sympathetic evaluation of deviant behavior.

The purpose of the present study was twofold: (a) to determine whether or not the knowledge of developmentally or historically relevant information about a person manifesting deviant behavior would tend to change one's perception of him, and (b) to extend Nunnally's findings by studying the attitudes toward individuals who are not labeled in any way, but who manifest varying degrees of behavioral disturbance.

METHOD

Subjects

The Ss consisted of 60 adult residents of Rochester, New York, who responded affirmatively to a telephone contact, in which it was explained that a survey was being conducted to study people's opinions about emotional disturbance. The Ss were told that they would be required to read some brief descriptions of people and answer some questions about them. If they agreed to participate, the questionnaire and a self-addressed, stamped return envelope were mailed to them. The names were randomly selected from males listed in the telephone book. This was done to avoid the potential non-random bias (e.g., unmarried, divorced, etc.) of females who are likely to have telephone listings. On the average, Ss used in the study were in their midforties and had completed slightly more than 13 yr. of education. Most of them were married and employed in jobs which were primarily skilled or professional in nature.

The Ss completed and returned one of two forms of the questionnaire. Thirty Ss were sent a questionnaire containing only a description of individuals, while the remaining 30 completed a questionnaire in which historically relevant material was added to the description of the behavior. Both questionnaires are described in detail later. Five Ss in each of the two conditions failed to complete the semantic differential ratings, although they did respond to other portions of the questionnaire. Consequently, some of the analyses were based on Ns of 30 in each condition, while in others, the N was 25.

Measures

The questionnaires consisted of an explanatory introduction—asking S to give his reactions to "descriptions of various people"—followed by six case vignettes, each describing the behavior of a different person. The vignettes were shortened versions of case histories found in Zax and Stricker's (1963) collection and were chosen from a pool of 17 histories which had been rated by eight clinical psychologists on a 7-point scale of degree of disturbance. The six vignettes used were the most representative of the two descriptions at each of the high, medium, and low degrees of disturbance.

In order to study the effect of historically relevant material on one's reaction to deviant behavior, two forms of the questionnaire were used. One form contained vignettes having only the description of the deviant behavior of particular individuals; this is referred to as Form D. The other form, in addition to providing the description of the behavior, also included material which was historically relevant to the behavior described; this is called Form D-H.

Following each vignette, the questionnaire contained a series of nine semantic differential scales, selected to represent an evaluative factor (Osgood, Suci, & Tannenbaum, 1957) and an understandability factor (Nunnally, 1961). The scales used for the

evaluative factor included effective-ineffective, good-bad, pleasant-unpleasant, sincere-insincere; for the understandability factor, the scales were familiar-strange, predictable-unpredictable, simple-complicated, understandable-mysterious, and straight-twisted. In addition, two previously used (Nunnally, 1961) and relevant scales (safe-dangerous and healthy-sick) were included, even though their factor loadings were indeterminate. The Ss were also asked to rate each individual described as to personal liking and degree of disturbance, as well as to judge whether the person had "some sort of mental illness."³

RESULTS

The semantic differential ratings, as well as the ratings of likeability and disturbance for the two vignettes at each level of disturbance, were averaged. Hence, the data for each S consisted of a series of ratings at each of the low, medium, and high levels of disturbance. The study employed a 2×3 Type I "mixed" design (Lindquist, 1953) where the effect of historically relevant material (D versus D-H) as well as level of disturbance (low versus medium versus high) on Ss' evaluation of deviant behavior were studied.

The results of the analysis of variance for the semantic differential ratings revealed that the historically relevant information provided for each case significantly improved Ss understanding of the deviant individuals described in the vignettes ($F = 7.08$, $df = 1/48$, $p < .05$). The background information, on the other hand, did not affect Ss' positive evaluation of the individuals described, nor did it produce any changes on the safe-dangerous or healthy-sick scales.

The level of disturbance reflected in each vignette had a striking effect on the semantic differential ratings. Regardless of whether or not background information was provided, the findings clearly point to the fact that with increasing levels of disturbance, the indi-

viduals described in the vignettes were significantly less understandable ($F = 138.35$, $df = 2/96$, $p < .001$), evaluated less positively ($F = 86.39$, $df = 2/96$, $p < .001$), and considered more dangerous than safe ($F = 137.22$, $df = 2/29$, $p < .001$), and more sick than healthy ($F = 97.49$, $df = 2/29$, $p < .001$).

The analysis of Ss' ratings of disturbance and likeability indicated that although the historically relevant background information did not affect these ratings, level of disturbance did. Hence, the more deviant individuals described in vignettes were seen by Ss as being both more severely disturbed ($F = 271.29$, $df = 2/116$, $p < .001$) and personally less likeable ($F = 89.24$, $df = 2/116$, $p < .001$).

Following each vignette, S was asked whether or not the individual described had "some sort of mental illness." The D and D-H groups were compared with respect to their response to this question by means of six chi-square analyses, one for each of the vignettes. Five of these comparisons failed to reach statistical significance. A significant chi-square was obtained with a vignette representing a medium level of disturbance, indicating that the availability of historically relevant information made it less likely that the individual described would be judge as being "mentally ill" ($\chi^2 = 5.93$, $df = 1$, $p < .02$).

DISCUSSION

The results of this study indicate that the availability of certain information, in this case, developmentally relevant material, increases the public's understanding or comprehension of displayed deviant behavior, while leaving unmodified an essentially negative evaluation of that behavior. In a sense, the present study succeeded in rendering more comprehensible what continues to be viewed as bad, unpleasant, ineffective behavior.

When presented with the descriptive vignette, each of which represented one of three predetermined (by clinician's ratings) levels of disturbance, Ss were able to differentiate accurately the behavioral descriptions according to their own estimates of level of disturbance. Regardless of whether or not develop-

³ A description of the vignettes appearing in Forms D and D-H, the rating scales of personal liking and degree of disturbance, and the analysis of variance summary tables may be obtained from the senior author or from the National Auxiliary Publication Service, Order Document No. 00550 from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Remit in advance \$3.00 for photocopies or \$1.00 for microfilm and make checks payable to: Research and Microfilm Publications, Inc.

mentally relevant material had been provided, however, Ss were less understanding of and reacted with a uniformly negative reaction to those individuals manifesting higher levels of disturbance.

The investigators had originally expected that inasmuch as mental illness is typically viewed as a disorder whose cause is unknown or vague, the existence of mental illness might be ruled out in those instances in which there was some understanding of the reasons for the deviant behavior. However, only the case of one vignette—a behavioral description representing a medium level of disturbance—was the developmentally relevant information able to minimize the tendency to apply the mental illness label. An intriguing, although admittedly preliminary explanation, may reside in a possible interaction between the level of disturbance and information on developmental antecedents, such that the attribution of mental illness to an individual displaying deviant behavior may be modified more readily in those instances where the behavior is only *moderately deviant*.

It might be emphasized that this study has attempted to alter the relationship between the deviance of certain behaviors, various judgments of that behavior, and attitudes toward the person who is engaging in these behaviors. Although the experimental manipulation was successful in providing a better understanding of the deviant behavior, the negative evaluations of the individuals were unchanged. It is possible that a more thorough understanding of the behavior is needed to bring about a more sympathetic and positive evaluation. Thus, a stronger treatment effect—involving more detailed information about

developmental history and/or information about concurrent situations which may have elicited the deviant behavior—would appear to be worth pursuing in future research.

REFERENCES

- ALTROCCHI, J., & EISDORFER, C. Changes in attitudes toward mental illness. *Mental Hygiene*, 1961, 45, 563-570.
- CUMMING, E., & CUMMING, J. *Closed ranks*. Cambridge, Mass.: Harvard University Press, 1957.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MAHER, B. A. *Principles of psychopathology*. New York: McGraw-Hill, 1966.
- NUNNALLY, J. C., JR. *Popular conceptions of mental health*. New York: Holt, Rinehart & Winston, 1961.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- ROTHAUS, P., HANSON, P. G., CLEVELAND, S. E., & JOHNSON, D. L. Describing psychiatric hospitalization: A dilemma. *American Psychologist*, 1963, 18, 85-89.
- SARBIN, T. R. On the futility of the proposition that some people be labeled "mentally ill." *Journal of Consulting Psychology*, 1967, 31, 447-453.
- STARR, S. A. The public's ideas about mental illness. Paper presented at the meeting of the National Association for Mental Health, Indianapolis, November 1955.
- SZASZ, T. S. *The myth of mental illness*. New York: Hoeber, 1961.
- ULLMANN, L. P., & KRASNER, L. *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.
- YARROW, M. R., SCHWARTZ, C. G., MURPHY, H. S., & DEASEY, L. C. The psychological meaning of mental illness in the family. *Journal of Social Issues*, 1955, 11, 12-24.
- ZAX, M., & STRICKER, G. *Patterns of psychopathology*. New York: Macmillan, 1963.

(Received May 31, 1968)

ATTITUDES TO BODY PRODUCTS AMONG NORMAL SUBJECTS¹

MICHAEL HIRT²

Kent State University

W. DONALD ROSS, RICHARD KURTZ,³ AND GOLDINE C. GLESER

University of Cincinnati College of Medicine

Attitudes of normal Ss toward body products were measured with the semantic differential. Ss of various ages, from different educational and socioeconomic backgrounds, were studied. Sex and race differences were also investigated. The results indicated that each group differentiated these body products significantly, and that the scaling of these products was consistent, and did not vary among the groups more than could have been expected by chance. The ordering of these body products and the absolute scale values assigned them were consistent. It was concluded that the data supported Kubie's theoretical paper which anticipated the universality of such attitudes and their hierarchial ordering.

The importance of attitudes toward the body has been studied in psychosexual development (Federn, 1952; Freud, 1931, 1933) among psychotic patients (Fisher, 1958; Witkin, 1962) and with healthy Ss (Traub & Orbach, 1964; Wapner & Werner, 1957). There are no empirical data, however, reporting the attitudes of healthy Ss toward specific body products such as excreta, tears, saliva, blood, etc. Kubie's (1937) "Fantasy of Dirt" seems to be the only paper which discusses such attitudes and is based upon clinical psychoanalytic study.

The purpose of this paper is to examine empirically the attitudes of normal Ss toward body products.

METHOD

Attitudes were measured with the semantic differential (Osgood, Suci, & Tannenbaum, 1957). The dimensions of evaluation, potency, and activity were sampled by three bipolar scales. Each S responded to 21 body products which, together with the form of the semantic differential used, are listed in Figure 1.

¹ This work was supported in part by United States Public Health Service Grants 5T1 MH-8018 and 5-S01-FR-5408-5. The authors wish to thank E. F. Buyniski, John Terlesky, Jan Robinson, and Mary Ann Hertz for their assistance in collecting these data.

² Requests for reprints should be sent to Michael Hirt, Department of Psychology, Kent State University, Kent, Ohio 44240.

³ Now at Washington University, St. Louis.

The results from a pilot study suggested that a single factor⁴ accounted for most of the communal variance (Harmon, 1960). Additionally, the scaling of these body products was found to be highly reproducible with generalizability coefficients in the .90s (Cronbach, Rajaratnam, & Gleser, 1963). It was therefore decided to explore the attitudes toward body products of normal Ss representing wider age, educational, and socioeconomic levels. The Ss were unselected job applicants at a large factory. The age range of the 205 Ss was 20-62 yr.; their educational level ranged 8-16 yr.; each of the five social classes described by Hollingshead and Redlich (1958) was represented.

This sample was divided into the following subgroups: six groups of relatively equal size of Caucasian males, below or above the age of 30, belonging to either a "high," "low," or "middle" social class; 39 Caucasian female Ss who were compared with a group of male Ss individually matched for age, education, and social class; and a group of 11 Negro Ss who were compared with a group of individually matched Caucasian Ss. The data were analyzed to bear upon the following two questions:

1. Is there, within any given subgroup, a consistent scale value assigned to these body products? In other words, are Ss consistent in the ordering of these body products and in the absolute scores ascribed to them?

2. Is the ordering and the absolute scores assigned these body products consistent among different subgroups, so that a single set of score values can be taken as representative of the attitudes of these body products for the population in general?

⁴ The score for the evaluative dimension plus the score for the activity dimension minus the score for the potency dimension.

Concept

Fresh*	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Stale
Thin**	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Thick
Fast***	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Slow
Sour*	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Sweet
Hard**	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Soft
Dull***	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Sharp
Clean*	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Dirty
Light**	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Heavy
Hot***	_____:	_____:	_____:	_____:	_____:	_____:	_____:	Cold

* Evaluative dimension

** Potency dimension

*** Activity dimension

Ear wax	Dandruff	Tears
Desquamated cells between the toes	Face Sweat	Pus
Nose pickings	Saliva	Phlegm
Hair	Blood	Nail clippings
Feces	Vomit	Pubic hair
Semen	Menstrual discharge	Body sweat
Human's milk	Blackheads	Urine

FIG. 1. Form and order of the semantic differential scales and list of body product concepts.

RESULTS

The differentiation of these body products was very high for each subgroup. The data for each subgroup were analyzed by means of analysis of variance, and the p values were, without exception, less than .001. An additional indication of the consistent scaling be-

havior of each subgroup was given by the estimated generalizability of the scores given by a single member of a given subgroup to the scores of the population that the subgroup represented. These reliability coefficients ranged from .37-.53 for a single individual. The equivalent index of the extent to which the obtained scaling could be repro-

TABLE 1
AVERAGE BODY PRODUCT SCORES BY DIFFERENT GROUPS

Body products	Groups							Females (N = 34)	Negroes (N = 11)
	Caucasian males								
	Age 20-30			Above age 30					
	Social class			Social class					
	I + II + III (N = 28)	IV (N = 27)	V (N = 33)	I + II (N = 24)	III (N = 20)	IV + V (N = 28)			
Tears	5.16	4.95	5.04	5.43	5.08	5.01	5.31	4.92	
Human's milk	5.24	5.08	5.02	5.31	4.95	4.97	5.17	4.93	
Blood	4.74	4.90	4.78	4.88	4.87	4.87	4.80	4.58	
Hair	4.71	4.76	4.70	4.62	4.77	4.50	4.97	4.50	
Saliva	4.19	4.38	4.45	4.65	4.45	4.12	4.49	4.24	
Semen	4.65	4.30	4.28	4.54	4.26	4.02	4.42	4.14	
Face sweat	4.23	3.88	4.25	4.40	4.05	4.08	4.24	4.19	
Urine	4.46	4.05	4.02	4.29	4.12	4.06	4.55	4.09	
Pubic hair	3.92	3.96	3.91	3.97	3.85	3.78	4.11	3.91	
Body sweat	4.04	3.76	3.72	4.03	3.80	3.88	3.99	4.04	
Menstrual discharge	3.64	3.76	3.47	3.82	3.60	3.60	3.95	3.57	
Desquamated cells									
Between the toes	3.29	3.26	3.68	3.62	3.51	3.54	3.32	3.34	
Feces	3.43	3.57	3.42	3.29	3.28	3.86	3.50	3.67	
Phlegm	3.37	3.48	3.50	3.35	3.23	3.35	3.39	3.76	
Nail clippings	3.34	3.30	3.34	3.44	3.46	3.35	3.43	3.58	
Dandruff	3.21	3.31	3.42	3.59	3.28	3.51	3.30	3.22	
Ear wax	3.30	3.21	3.36	3.39	3.38	3.28	3.34	3.52	
Vomit	3.04	3.19	3.00	3.45	3.28	3.36	3.30	3.09	
Nose pickings	3.40	2.95	3.24	3.33	3.08	3.32	3.24	3.31	
Pus	2.98	3.18	3.14	3.11	3.08	2.97	3.23	3.25	
Blackheads	2.91	2.72	2.90	3.10	2.85	3.09	3.06	2.90	

Note.—I-V = the five social classes described by Hollingshead and Redlich (1958).

duced by other comparable subgroups ranged .90-.97.

The comparability of the scale values assigned these body products by each of the subgroups was determined by comparing the variability of assigned scale values between the subgroups with the variability within each of the subgroups. In no case was the variability between subgroups greater than could have been expected from sampling fluctuations. The implication of this result is that the intraclass correlations among these body product concept scores, at least for the population(s) sampled, would be perfect (1.00). Since no differences were found between these subgroups, they may be assumed to come from the same population, providing at least partial support for the "universality" of attitudes toward these body products.

In Table 1, the body products have been ordered (from high to low) on the basis of

their score from all Ss. The score assigned each body product by each subgroup is given in original semantic differential units. It is readily apparent that both the ordering and the absolute scores for each body product remained markedly consistent among the different groups. Tears and human's milk always received the highest score, while pus and blackheads were always among the lowest ranked body products. The other body products also retain their ranks, with occasional movement up (or down) one position in the order.

An additional observation from Table 1 is the consistent use by all Ss of the lower end of the semantic differential scale. In other words, although the semantic differential permits a score from one to seven, most Ss used only the lower end of the scale, thus expressing negative attitudes toward most body products.

DISCUSSION

The major finding of this study is that there does indeed exist a consistent hierarchy of attitudes toward body products. Kubie's theoretical paper anticipated correctly the universality of such attitudes and their hierarchical ordering. The high correlation which has been reported between evaluative factor loadings and social desirability ratings of semantic differential scales (Ford & Meisels, 1965) lends further support to Kubie's formulation.

The meaning of this hierarchy of body products is still relatively obscure. One possibility is that these body products were ordered along a body-alien to body-syntonic continuum. In other words, body products may be perceived as "normal" components of on-going body functions (e.g., blood, tears, saliva) or may be discarded or somehow sloughed off (e.g., nail clippings). The identification of the more specific properties of body products (e.g., their odor, color, consistency, etc.) which are of major significance remains as an area for further investigation.

REFERENCES

- CRONBACH, L. J., RAJARATNAM, N., & GLEESER, G. C. Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 1963, 16, 137-163.

- FEDERN, P. *Ego psychology and the psychosis*. New York: Basic Books, 1952.
- FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton: Van Nostrand, 1958.
- FORD, L. H., JR., & MEISELS, M. Social desirability and the semantic differential. *Educational and Psychological Measurement*, 1965, 25, 465-475.
- FREUD, S. Female sexuality. In, *Collected papers*. Vol. 5. New York: Basic Books, 1931.
- FREUD, S. The psychology of women. In, *New introductory lectures on psychoanalysis*. New York: Norton, 1933.
- HARMAN, H. *Modern factor analysis*. Chicago: University of Chicago Press, 1960.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- KUBIE, L. S. The fantasy of dirt. *Psychoanalytic Quarterly*, 1937, 6, 388-424.
- OSGOOD, C., SUCH, G., & TANNENBAUM, P. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- TRAUB, A. C., & ORBACH, J. Psychophysical studies of body-image. *Archives of General Psychiatry*, 1964, 11, 53-66.
- WAPNER, S., & WERNER, H. *Perceptual development*. Worcester, Mass.: Clark University Press, 1957.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation*. New York: Wiley, 1962.

(Received June 10, 1968)

CERTAIN INITIAL VARIABLES AS PREDICTORS OF CHANGE WITH CLASSICAL PSYCHOANALYSIS¹

H. ARONSON²

AND

WALTER WEINTRAUB

*Department of Psychiatry,
Temple University Health Sciences Center*

*The Psychiatric Institute,
University of Maryland School of Medicine*

The full case loads of 28 analysts were studied to determine predictors of change with psychoanalysis. Information was obtained by questionnaire from the treating analysts. It was found that, while initial complaints are of little value in predicting success in analysis, they do relate to dropout from treatment. Motivational factors such as breadth and degree of discomfort, external stress, and chronicity of disorder relate to both success in analysis and dropout from treatment. It was concluded that, as long as symptoms are not extremely pathological, improvement in and commitment to analysis is determined by the pressure the disorder places on the patient.

Classical psychoanalysis is a form of therapy that has had, despite a long period of widespread interest and attention, little experimental scrutiny. As a result, many beliefs and impressions have become established over years of application; yet, few controlled studies exist to provide even the simplest requisite data upon which to base experimental validation.

Since analysis demands a relationship with considerable commitment on the part of both patient and therapist, it is difficult to manipulate variables externally. Those interested in research on this form of psychotherapy must, therefore, often make a difficult choice. They must either rely upon lengthy interpolation of laboratory results or use epidemiological methods and work without the safeguards of random procedures or controls. We have chosen the second alternative.

People screened as being suitable for analysis are far from a random group of applicants for psychotherapy. Should it be possible, within the rather narrow range of patients already acceptable, to specify variables that

predict greater or lesser improvement, those predictive variables would have both theoretical and practical value.

In other reports we detailed our method of data collection and tabulated the descriptive (Weintraub & Aronson, 1968) and social (Aronson & Weintraub, 1968b) backgrounds of patients who were the Ss of our investigation. Then we began to examine the changes that could be measured as occurring in the course of analytic treatment (Aronson & Weintraub, 1968a). The present paper will focus upon the predictive value of initial variables. Our goal is to determine, using information available early in analysis, those patients who could be expected to improve as well as to distinguish those who are likely to remain from those who are apt to leave analysis short of completion.

METHOD

Procedure

It is important to note that our work has been based upon the assumption that groups of patients, each studied at a different longitudinal point within analysis, would approximate the changes that might occur in a hypothetical individual as he progressed through his analytic treatment. Thus, while we measure discrete intervals of analysis, we will discuss our findings in terms of curves that indicate the effects of time.

Our data indicate a highly significant correlation between improvement on each of the scales chosen to measure change and duration of analysis. As can be seen in Figure 1, a hypothetical curve of improvement would be almost identical for both those patients who remained in analysis and those who left analysis short of completion. Figure 2 indicates

¹ This investigation was supported in part by United States Public Health Service Award K3-MH-19393, and Grant MH-07434, from the National Institutes of Health, United States Public Health Service.

The authors wish to thank Jean Warthan for her generous aid in programming for computer analysis some of the data presented here.

² Requests for reprints should be sent to Harriet Aronson, Department of Psychiatry, Temple University Health Sciences Center, Philadelphia, Pennsylvania 19140.

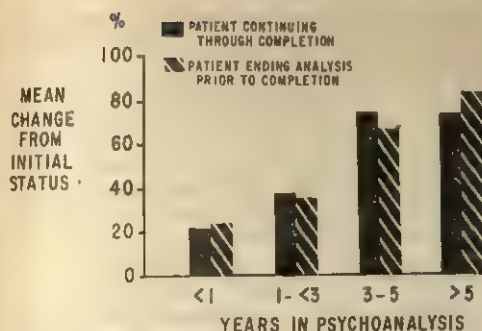


FIG. 1. Mean achievement of potential improvement in patients continuing through completion of psychoanalysis and in patients leaving analysis short of completion.

however, that those patients who left prior to completion included a majority of those who became worse. Our hypothetical improvement curve seems to take a steady (learning) rather than a sudden (insight) incline with the greatest increment in the period between the second and fourth years. Our focus here will center upon those variables that may relate to greater learning with analysis.

In examining the predictive value of initial variables, we found the most obvious, that is, diagnosis, to have little relation to the rate of improvement (Aronson & Weintraub, 1968a). However, because diagnoses are rough estimates, difficult to establish with reliability, it appeared reasonable to use more refined means to designate initial status. This became the basis of our continuing investigation. The variables that will be examined here are: the kind and number of presenting symptoms and/or complaints, the chronicity of such symptoms, the degree of subjective discomfort, and the presence of external stress.

While no specific hypotheses were chosen prior to obtaining the data, a number of questions can be posed:

1. What kinds of symptoms or complaints do analytic patients present?
2. Is there a relationship between the kind or variety of initial symptoms presented by a patient and the results of his analysis?
3. Does the degree of the patient's discomfort relate to improvement?
4. Does it matter whether the impetus for treatment is internal or external (environmental pressure)?
5. Can anything be stated as to why patients leave analysis?
6. What determines the point at which dropout from analysis occurs?

Subjects

The data reported here were collected during the course of a prevalence survey of patients undergoing analysis during the then current "psychoanalytic

year" (September 1962 through August 1963). Thirty analysts, approached informally, agreed to supply information on *all* patients they had seen in classical psychoanalysis during the studied year. Our sample covers the total analytic case loads of this non-randomly selected group of analysts with the following exceptions: (a) Data concerning several patients, known by their analysts to be personally acquainted with the authors, were excluded. (b) Two analysts, with a combined total of 17 patients, responded to a shorter set of selected questions. The variables studied here were not covered in the information provided by these two analysts.

The sample for this report thus includes a total of 127 cases reported upon by 28 analysts. The information collected relates to both the initial condition of the 127 patients and to their condition during the surveyed period. As a prevalence study, the information obtained represents patients seen at all points in their analyses from initiation to either mutually determined completion or a unilateral decision to discontinue treatment.

Measurement of Change

Due to the nature of our survey, all measures were limited to judgments of the reporting analysts. To increase reliability, fixed-point scales were used as our method of choice. Unfortunately, it was not possible to construct an acceptable fixed-point scale to measure symptomatic change, and a scale that covered four degrees of change was used instead. These were: (1) complete remission or greatly improved; (2) moderately improved; (3) minimal improvement; (4) exacerbation of symptoms. Fixed-point scales were constructed to measure: vocational adjustment; object relations (relationships with others); and capacity to experience pleasure. Vocational adjustment, shown as an example below, covered three steps; the remaining scales covered four steps:

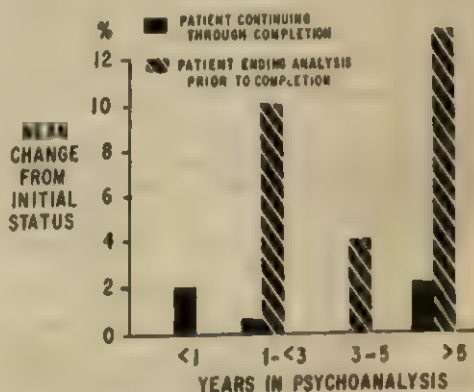


FIG. 2. Mean occurrence of potential regression in patients continuing through completion of psychoanalysis and in patients leaving analysis prior to completion.

1. Patient unable to find work or keep a job (for housewife: unable to perform responsibilities of wife and mother; for student: unable to make passing grades).

2. Patient able to keep position but not at a level appropriate to his intelligence, training, and experience (for housewife: able to perform responsibilities of wife and mother, but at a low level of efficiency; for student: able to make passing grades, but not performing up to potential).

3. Patient able to work at level appropriate to intelligence, training, and experience (for housewife: carries out responsibilities of wife and mother efficiently; for student: performance in school is up to potential).

The four scales of symptomatic status and level of

functioning, described more fully in Aronson and Weintraub (1968a), were our criteria of change with analysis.

It is important to note that a patient initially rated at the extreme of a fixed-point scale cannot be measured as moving farther in the direction of his extreme rating. Our results, therefore, relate changes only for those patients who were changeable in the direction being studied. Our *N*s vary accordingly for each of the four measures.

Adjustment for Duration of Analysis

By chance or due to differential drop-out rates, patients with a particular complaint or condition might easily be overrepresented in the earlier years of analysis. Their mean improvement rate, as com-

TABLE 1

INITIAL SYMPTOMS OR COMPLAINTS, ARRANGED INTO MAJOR CATEGORIES, OF 126 PSYCHOANALYSANDS

Category	No. patients	Category	No. patients
Neurotic		Psychosomatic (continued)	
Obsessions and/or compulsions	50	Menstrual disturbances	4
Phobias	36	Migraine	3
Boredom	27	Asthma	3
Conversion symptom(s)	22	Peptic ulcer	1
Social inhibitions	4	Paroxysmal tachycardia	1
Dissociative episode(s)	2	Overweight	1
Work or study inhibitions	1	Dermatitis	1
Anxiety over sexual identity	1	Insomnia	1
At least one symptom in category	94 (75%)	At least one symptom in category	37 (29%)
No other category listing	5	No other category listing	1
Habits		Character traits	
Nail-biting	7	Gross immaturity	19
Tic(s)	2	Hypochondriasis	12
Enuresis	2	Withdrawn or asocial	9
Encopresis	1	Narcissism	1
Bite cheek	1	Masochism	1
At least one symptom in category	10 (8%)	At least one symptom in category	35 (28%)
No other category listing	0	No other category listing	1
Alloplastic		Sexual inhibition or inadequacy	
"Acting-out" behavior	45	Impotence or frigidity	48
Extremely low frustration tolerance	29	Premature ejaculation	2
Temper tantrums	21	Sexual inhibitions	1
Alcoholism	4	At least one symptom in category	51 (40%)
Gross antisocial behavior	3	No other category listing	1
Infidelity	1	Sexual deviation	
Truancy	1	Homosexuality	16
At least one symptom in category	64 (51%)	Voyeurism	6
No other category listing	4	Pedophilia	2
Severe anxiety		Fetishism	2
Total	67 (53%)	Exhibitionism	2
No other category listing	0	Sexual sadism	2
Depression		Transvestism	1
Total	59 (47%)	Kleptomania	1
No other category listing	1	At least one symptom in category	25 (20%)
Psychosomatic		No other category listing	1
Chronic fatigue	16	Loss of reality	
Headaches	10	Paranoid	3
Backaches	6	Depersonalization	2
Colitis	5	At least one symptom in category	4 (3%)
		No other category listing	0

pared with patients who continue for further treatment, would then be deceptively low. In order to adjust for variability due to unequal time in analysis, the data reported upon here were, wherever appropriate, divided between those from patients in analysis for less than 3 yr. and those in analysis for 3 yr. or more. The differences attributable to each variable being studied were then combined from both durational levels by means of Cochran's criterion for combining results from groups of unequal size (Maxwell, 1961). This method essentially approaches the data as if we had done two studies, one of a group of patients seen early in treatment and another of patients seen late in analysis. The results of the two are combined to use the larger *N* while still controlling for durational effects.

Since no directional hypotheses were made, two-tailed tests of significance were used in determining significance levels.

RESULTS

Category of Initial Complaints

A list of 48 prominent symptoms with blank spaces for additional unlisted symptoms or complaints provided each analyst a framework upon which to report the initial problems of each patient. Table 1 shows those symptoms or complaints present at initiation of analysis, each listed under 1 of 10 major categories. Symptoms were placed into categories without knowledge of other results and, wherever placement was in doubt, categorization followed analytic theory. One write-in symptom, "problems with children," could not be categorized, and that patient, who had no other symptom recorded, was dropped from those parts of the study requiring categorization of symptoms.

Categories were chosen as being more specific than diagnoses yet more encompassing than symptoms. A patient with two symptoms in a single category is treated as having only one symptom category. His pathology is considered as being narrower than that of a patient whose symptoms cross category lines. It is important to note that a patient can be listed under any or all of the categories simultaneously and, in fact, few patients had difficulties limited to only one category.

Data concerning the number of patients, by years in analysis, found under each category as well as under other variables to be discussed below are presented in Table 2. Table 3 lists the improvement rates by category of patients who had at least one symptom in the

TABLE 2
NUMBER OF PATIENTS STUDIED IN PREVALENCE SURVEY OF ANALYTIC TREATMENT, CATEGORIZED BY STATUS AT INITIATION OF ANALYSIS

Initial variable	Years in analysis	
	<3	≥3
Symptom category		
Neurotic	60	33
Habits	5	5
Alloplastic	38	25
Severe anxiety	44	23
Depression	37	22
Psychosomatic	25	11
Character traits	21	13
Sexual inhibition or inadequacy	32	19
Sexual deviation	15	10
Loss of reality	2	2
Number of symptom categories		
>6	10	6
3-5	45	28
<3	22	15
Degree of subjective suffering at initiation		
Extreme	22	15
Moderate	45	26
Mild	9	8
Treatment initiated after onset or exacerbation of acute symptomatology		
Yes	36	42
No	41	7
Treatment initiated after a major external situational stress		
Yes	15	16
No	62	33

category. The differences between rates for patients found within a symptom category and those obtained for patients not included under that category were compared. The resulting significances did not exceed the number expected by chance. It would appear that knowledge of initial symptoms or complaints is as little useful as was crude diagnosis in predicting improvement rates within this already screened patient group.

Motivational Effects

The findings on four variables, motivational in nature, are shown in Table 4. They were: breadth and depth of discomfort, chronicity of disorder, and the presence of external pressure. Breadth of discomfort was measured by the total number of different symptom categories under which a patient's initial com-

TABLE 3

ACHIEVEMENT OF POTENTIAL IMPROVEMENT OF ANALYTIC PATIENTS
CATEGORIZED BY INITIAL STATUS OF SYMPTOMS

Symptom category	Measure of improvement (by years in analysis)											
	Symptomatic			Vocational functioning			Object relationships			Capacity for pleasure		
	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>
Neurotic	50%	94%	<.05	28%	88%	<.10	23%	59%	<i>ns</i>	37%	81%	<i>ns</i>
Habits	60	60	<i>ns</i>	0	80	<i>ns</i>	0	60	<i>ns</i>	40	80	<i>ns</i>
Alloplastic	39	80	<.10	15	73	<.10	15	52	<i>ns</i>	36	70	<i>ns</i>
Severe anxiety	50	91	<i>ns</i>	25	87	<i>ns</i>	20	68	<i>ns</i>	34	83	<i>ns</i>
Depression	51	86	<i>ns</i>	32	77	<i>ns</i>	27	55	<i>ns</i>	35	70	<i>ns</i>
Psychosomatic	56	82	<i>ns</i>	35	71	<i>ns</i>	35	73	<.05	36	64	<i>ns</i>
Character traits	48	86	<i>ns</i>	44	89	<.10	30	50	<i>ns</i>	29	71	<i>ns</i>
Sexual inhibition or inadequacy	47	95	<i>ns</i>	23	88	<i>ns</i>	17	47	<i>ns</i>	28	84	<i>ns</i>
Sexual deviation	47	90	<i>ns</i>	33	67	<i>ns</i>	23	33	<i>ns</i>	47	70	<i>ns</i>
Loss of reality	50	100	<i>ns</i>	50	—	<.10	50	50	<i>ns</i>	50	100	<i>ns</i>
Total group	47	86		25	76		21	57		34	78	

* Each probability test measures the improvement of those patients within the symptom category as compared to all other patients. Differences were analyzed statistically by means of Cochran's criterion for combining results from groups of unequal size. Probabilities are stated for two-tailed conditions. By chance, 2 of these 40 comparisons should reach $p < .05$.

plaints were listed. Here a significant difference was found for three of our four measures of improvement. However, although those with greater breadth had greater early improvement, less regularity was found in the later therapeutic years. In general, while the

TABLE 4

ACHIEVEMENT OF POTENTIAL IMPROVEMENT OF ANALYTIC PATIENTS CATEGORIZED
BY INITIAL STATUS OF MOTIVATIONAL FACTORS

Initial status	Measure of improvement (by years in analysis)											
	Symptomatic			Vocational functioning			Object relationships			Capacity for pleasure		
	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>	<3	≥3	<i>p</i>
No. symptom categories												
>5	60	83		43	100		25	33		50	67	
3-5	44	89	<i>ns</i>	27	71	<.03	23	70	<.03	27	78	<.05
<3	45	80		14	78		15	38		41	85	
Subjective discomfort												
High	55	93		33	92		30	64		41	87	
Medium	44	85	<i>ns</i>	27	69	<.03	19	56	<.05	33	76	<.10
Low	44	75		14	50		12	43		22	67	
Acute onset												
Yes	59	83	<.10	21	93	<.10	25	65	<i>ns</i>	47	83	<.03
No	38	87		28	60		19	52		24	75	
External stress												
Yes	47	94	<i>ns</i>	10	100	<.03	8	60	<i>ns</i>	27	79	<i>ns</i>
No	45	82		29	65		24	55		35	78	
Total group	47%	86%		25%	76%		21%	57%		34%	78%	

* Differences were compared by Cochran's criterion for combining results from groups of unequal size. Probabilities are stated for two-tailed conditions.

breadth of difficulty appeared to be important in terms of statistical significance, its effect was more consistent early in analysis.

Such irregularity was not the case for the degree of subjective discomfort. With subjective discomfort there was a regular progression such that, both for those in the first 3 yr. and those later in analysis, the greater the discomfort the more likely the improvement. This progression reached statistical significance for two measures and approached significance ($p < .10$) for another. It was still present but insignificant for the fourth measure of improvement.

An examination of chronicity also produced statistically significant differences. The patients who, on initiating analysis, were in either acute or exacerbated states generally showed greater improvement. This was significant for one measure and showed a tendency toward significance ($p < .10$) for two more. The presence of external or situational stress had a significant effect on only one measure, vocational functioning. Such stress appeared to retard vocational improvement in the first 3 yr. and increase improvement during the later years of analysis, as compared with change found in patients lacking external stress.

Of the 16 statistical tests done on motivational effects, we should expect less than 1 to be significant by chance at the .05 level. The data showed 7 comparisons significant at the .05 level with an additional 3 at the .10 level. Unlike our findings with symptom categories, motivational variables appear to have effects that cannot be explained by chance alone.

Unilaterally Determined Termination

Table 5 records the percentages of patients who left analysis for reasons other than completion of treatment. Completion may be defined as that point at which analyst and patient agree that maximum benefit has been obtained from treatment. While there is not a significant difference in overall drop-out rates between durational periods, we have retained the division of up-to-3 and 3-or-more years in analysis. This division allows for greater interpretation in terms of reasons for premature termination.

TABLE 5

DROP-OUT RATE DURING SURVEY YEAR OF ANALYTIC PATIENTS CATEGORIZED BY SYMPTOMS AND OTHER INITIAL VARIABLES

Symptom or variable	% dropout		Probability of difference ^a
	< 3 yr. of analysis	≥ 3 yr. of analysis	
Symptom category			
Neurotic	12	18	ns
Habits	20	20	ns
Alloplastic	5	20	ns
Anxiety	9	13	ns
Depression	5	14	ns
Psychosomatic	4	0	<.05
Character traits	5	36	<.03
Sexual inhibition or inadequacy	12	16	ns
Sexual deviation	7	5	ns
Loss of reality	50	100	<.001
No. symptom categories			
≥ 6	10	50	
3-5	7	7	<.01
< 3	18	20	
Degree subjective suffering at initiation			
Extreme	9	27	
Moderate	9	12	<.10
Mild	22	12	
Exacerbation or acute symptoms prior to initiation			
Yes	6	17	ns
No	13	16	
Major external situational stress prior to initiation			
Yes	13	18	ns
No	10	15	
Total group	10	16	ns

^a Total group differences due to years in analysis were compared by the χ^2 test. The remaining statistical analyses for differences between groups give probabilities obtained with Cochran's criterion for combining results from groups of unequal size. All probabilities are stated for two-tailed conditions.

Here, the data do show differences attributable to symptom category. Patients whose symptoms included signs of distortion of reality consistently dropped out of treatment more often than did patients who did not show such initial complaints ($p < .001$). Patients with psychosomatic symptoms had exceptionally low drop-out rates ($p < .05$). For patients who had character traits, there was less dropout early in treatment and considerably more attrition later in treatment than for patients not listed under that category ($p < .03$).

Breadth of disorder showed a significant ($p < .01$) relationship with dropout. Those patients with a medium number of symptom categories showed the lowest attrition rate. Those with relatively few symptom categories had higher drop-out rates early in analysis while those with a wider range of symptoms had a disproportionately high percentage of dropouts late in analysis. Depth of disorder shows a similar trend. Those whose initial suffering was mild tended to drop out early while those whose suffering was extreme tended to leave prior to completion relatively late in treatment ($p < .10$).

There was no evidence that a major external situational stress influenced the rate of attrition.

DISCUSSION

Comparison with Other Studies

Strupp, Wallach, and Wogan (1964) studied a group composed of analytic as well as other patients in dynamic psychotherapy, for the most part without differentiating between depths of therapy. Their data indicate a close correspondence with our own findings on percentages of patients with such symptom categories as somatic problems, anxiety, psychoticlike signs, etc. Their patients included a noticeably smaller number with sexual problems and fewer of our group were listed as acting out (alloplastic). They obtained a significant correlation between their therapist's estimate of initial motivation and overall success in therapy, a finding consonant with our own. Unfortunately, they did not enlarge upon what they meant by motivation. Motivation may be defined in a number of ways. Wallerstein, Robbins, Sargeant, and Luborsky (1956), for example, defined motivation in terms of honesty, fees, goals, and secondary gains. These would not correspond to what we have considered as motivational matters.

Luborsky (1962), who reported interim results of the Menninger Project findings, concluded that for patients not greatly incapacitated at initiation, "the higher the level of anxiety the greater the change that is likely." Our group, as outpatients with sufficient "ego strength" to be acceptable for analysis, would generally fit the criterion of not being greatly incapacitated. However,

presence of severe anxiety, as our analysts reported it, did not predict either success or attrition from treatment.

The Central Fact-Gathering Committee of the American Psychoanalytic Association (Hamburg, 1967) reported a total of 43% of their patients as not completing analysis, a figure not unreasonably different from our attrition rate of 10+% per analytic year. If we were to reduce a theoretical 100 patients by the average drop-out rate found for each year, we would lose 52 by the fifth year.

The Fact-Gathering Report divided symptom categories into nine exclusive rather than overlapping areas, and often presented results on analytic patients and patients treated by psychotherapy without differentiating between treatments. It found low completion rates for schizophrenics, paranoids, and sexual deviates and a high completion rate for psychosomatic patients. Our findings are similar except for the low rate for sexual deviates. It is possible that, since our categories were not mutually exclusive, a number of less habituated cases were listed within our deviate category while the Fact-Gathering report may have included only cases for which a sexual problem was a primary concern.

In all, the differences in research methodology between our own and other studies were so great that even such comparisons as stated above are questionable.

Present Findings

It appeared, when we studied the diagnostic designations of our patients (Aronson & Weintraub, 1968a), that the majority of patients were labeled as character disorders. Since analysis was developed largely as a treatment for neurotics, this was an interesting result. However, an examination of actual symptoms would suggest that, if our results are typical of analytic practice, the problems being treated in analysis may still largely be neurotic ones. A majority of our patients had at least one neurotic complaint.

Our results bring us to conclude that either the kind of symptom presented is not especially important in predicting improvement, or that analysts are already well grounded in the methods they use to decide which symptomatic conditions can be improved by means

of their techniques. Our data indicate, for example, that patients whose contact with reality is not strong are seldom accepted for analysis; when accepted, they seldom complete the analytic course.

The major findings reported here indicate the importance of motivational factors both on improvement and on attrition from treatment. Patients who suffer more tend to benefit more from analysis both in the early and the later stages. Those whose symptoms range into a wider number of areas show particularly high rates of improvement early in treatment. Motivation by environmental stress is less effective in producing improvement than is the impetus of internal discomfort.

The patient most likely to benefit from analytic treatment is one whose symptoms are not so severe as to include distortion of reality. His discomfort is of relatively acute onset, giving him a recent memory of adequacy and less adaptation to disorder. His symptoms cover sufficient areas of life to make it difficult to avoid them. Yet he does not have so many symptoms that he has little or no experience with symptom-free functioning. His internal reaction to his lessened functioning is one of strong discomfort.

For this ideal patient, it is relatively unimportant to note through which specific symptom or symptoms his discomfort makes itself known. However, if his pathology includes character traits, he may not remain in analysis after some initial improvement has been obtained; if his symptoms include psychosomatic ones, he is more likely to be a tenacious patient.

Patients whose symptoms cover a narrow range or who have little personal distress are not only less likely to improve but also more likely to drop out of analysis during the first years of treatment. Of those who suffer extremely and of those whose symptoms cover a wider range, few drop out early in treatment; such patients, however, are more prone than less disabled patients to leave analysis following a unilateral decision after a number of years. By that time either motivational pressures have been lessened by improvement or their drive to change has led to further de-

terioration. In general, those who suffer more stay until some change has occurred.

The question of what happens to patients who leave analysis short of completion has seldom been broached. We have found that, at the point of departure, their achievement of positive change could not be differentiated from that of patients with equal therapeutic experience who remained for further analytic examination. Cases reported upon by Freud are remarkable, by present-day standards, for their brevity. Perhaps Freud ended his treatment with the expectation that the patient would continue to improve once the process was set in motion. Perhaps he counted on the motivations noted here to bring about the necessary changes. Present-day analysts, instead, direct the ongoing changes, thereby increasing treatment time. Apparently they hold continuing motivation as insufficient to achieve the desired effects. No experimental studies have been reported that test the above contending hypotheses.

REFERENCES

- ARONSON, H., & WEINTRAUB, W. Patient changes during classical psychoanalysis as a function of initial status and duration of treatment. *Psychiatry*, 1968, 31, 369-379. (a)
- ARONSON, H., & WEINTRAUB, W. Social background of the patient in classical psychoanalysis. *Journal of Nervous and Mental Disease*, 1968, 146, 91-97. (b)
- HAMBURG, D. A. *Report of ad hoc committee on central fact-gathering data*. New York: American Psychoanalytic Association, 1967.
- LUBORSKY, L. The patient's personality and psychotherapeutic change. In H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy*. Baltimore: French-Bray, 1962.
- MAXWELL, A. E. *Analyzing qualitative data*. New York: Barnes & Noble, 1961.
- STRUPP, H. H., WALLACE, M. S., & WOGAN, M. Psychotherapy experience in retrospect: Questionnaire survey of former patients and their therapists. *Psychological Monographs*, 1964, 78(11, Whole No. 588).
- WALLERSTEIN, R. S., ROBBINS, L. L., SARGEANT, H. D., & LUBORSKY, L. The psychotherapy research project of the Menninger Foundation. Rationale, method and sample use. *Bulletin of the Menninger Clinic*, 1956, 20, 221-250.
- WEINTRAUB, W., & ARONSON, H. A survey of patients in classical psychoanalysis: Some vital statistics. *Journal of Nervous and Mental Disease*, 1968, 146, 98-102.

(Received June 24, 1968)

EFFECTS OF A "REFORMED SINNER" AND A "LAPSED SAINT" STRATEGY UPON TRUST FORMATION IN PARANOID AND NONPARANOID SCHIZOPHRENIC PATIENTS¹

THOMAS HARFORD

AND

LEONARD SOLOMON²

Massachusetts General Hospital, Boston

Boston University

Paranoid and nonparanoid schizophrenic patients, as well as a normal sample, were compared on: (a) capacity to form a trusting relationship with an unknown other, and (b) the relative efficacy of a "reformed sinner" versus a "lapsed saint" strategy in inducing trusting and trustworthy behavior in these samples. The Ss played a two-person bargaining game for 30 trials and were randomly assigned to one of two experimental strategies. Several game-playing styles differentiated the three samples. The paranoids, as contrasted with the nonparanoids and normals, were more likely to be "cooperative" when dealing with a lapsed saint and more "exploitative" when confronted with a reformed sinner. The results reflected the paranoids' conflicting attitudes toward discrepancies in power.

Several psychoanalytic theorists have proposed that the primary deficit in schizophrenia is an impairment in "basic trust" (Arieti, 1961; Erikson, 1953; Fromm-Reichmann, 1948). While various subtypes of schizophrenia are characterized by ambivalence with respect to when and whom to trust, the paranoid shows pervasive projection and rigid interpersonal suspicion (Wiener, 1966). Several factor-analytic studies by Lorr (1955, 1960, 1966) indicate that all types of paranoid patients show, in common, a reliance on projection, perceptual distortion, and overt hostility. Lorr (1966) defines projection as the "fixed attribution of hostile persecuting intent," whereas distortion refers to "hallucinations that threaten or demand [p. 35]."

The current practice of dichotomizing schizophrenia into paranoid and nonparanoid types (Katz, Cole, & Lowry, 1964) is consonant with research findings in the area of cognitive, perceptual, and psychomotor performance (Lynn, 1963; McGonaghy, 1960; Shakow, 1963; Silverman, 1964). Unfortu-

nately, few laboratory experiments have assessed the interpersonal motives and social emotional orientation of the paranoid schizophrenic. The present study seeks to determine how interpersonal trust and/or suspicion develops and how they can be modified in a sample of paranoid and nonparanoid schizophrenic patients.

Deutsch (1958, 1962) introduced the use of a two-person, bargaining game for the illumination of some of the personal and situational determinants of interpersonal trust and suspicion.³ Interpersonal trust is said to exist when a person (P) behaves on the expectation that the other person (O) will do something, and P perceives that O is aware that his behavior has motivational relevance for P.⁴

³ The essential ingredients of a bargaining relationship include (a) a conflict of interests between the participants, (b) provision of more than one outcome (i.e., a bargain may be struck or not), and (c) the possibility of cooperative strategies.

⁴ Trust refers to expectations of benevolence, whereas suspicion refers to expectations of malevolence. Thus, for example, if a car owner trusts a mechanic to repair his brakes properly and the mechanic violates his trust, the owner may suffer major harm—an auto accident. The trusting owner, when his trust is fulfilled, is in a more advantageous position than if he had not trusted. However, the disadvantages of trusting and having his trust unfulfilled are considerably greater (i.e., auto accident) than the advantages of having his trust fulfilled (i.e., meet appointments).

¹ This present report is based in part on a doctoral dissertation submitted by the first author to the graduate faculty, Boston University. This study was supported in part by Public Health Service Research Grant M-5353 from the National Institute of Mental Health.

² Requests for reprints should be sent to Thomas Harford, Massachusetts General Hospital, Boston, Massachusetts 02215.

		PERSON 1 (CONFEDERATE)	
		RED	GREEN
PERSON 1 (SUBJECT)	RED	+ 3 + 3	+ 5 - 5
	GREEN	- 5 + 5	- 3 - 3

FIG. 1. Game matrix. (Person 1 chooses a row; Person 2 chooses a column. Person 1's payoffs are in the lower left hand corner of each cell.)

Figure 1 presents the payoff matrix for the two-person bargaining game, the "prisoner's dilemma." In this game, interpersonal trust is defined as a cooperative choice based on an expectation of cooperation from the other player, whereas suspicion is defined as a non-cooperative choice. The paradoxical feature of the prisoner's dilemma is that, from the individual player's point of view, rational behavior is not possible unless a relationship of mutual trust exists.

Klein and Solomon (1966) studied a sample of acute schizophrenic patients in a two-person bargaining game. Each patient was paired with an experimental confederate who assumed an "exploitative" strategy for the first 15 trials. In the second phase the confederate "reformed" and adopted a conditionally cooperative strategy. The results indicated that the paranoids maintained their noncooperative stance for a longer period in the face of the confederate's reform. Klein and Solomon (1966, p. 507) questioned whether the "sinner's" initial exploitation served to reinforce the paranoid's suspicion (thereby making them more impervious to a cooperative reform), or whether this rigidity reflected a cognitive inability to adapt to change. To study these issues it would be necessary to design an experiment in which the confederate's strategy shifts from unconditional to

conditional cooperation ("lapsed saint") as well as from unconditional noncooperation to conditional cooperation ("reformed sinner"). Studies by Bixenstone and Wilson (1963) and Scodel (1959) with college students demonstrated that a reformed-sinner strategy is more effective in eliciting cooperation than such strategies as initial cooperation followed by noncooperation (lapsed saint) or unconditional cooperation (turn the other cheek).

A study by Knapp and Podell (1968) compared the behaviors of college students, psychotic patients, and prisoners in a two-person game in which the other player, a confederate, adopted either a cooperative strategy (80% cooperation) or a competitive strategy (20% cooperation). Although the competitive program failed to differentiate the three groups, the students and patients responded with more cooperation to the cooperative program, in contrast to the prisoners.

The present experiment sought to compare paranoid with nonparanoid schizophrenic patients, and a normal sample on (a) capacity to form a trusting relationship with an unknown other and (b) relative efficacy of a reformed-sinner versus a lapsed-saint strategy in inducing trusting and trustworthy behavior.

METHOD

Thirty-seven hospitalized schizophrenic patients from the Massachusetts Mental Health Center, and

11 patients from McLean Hospital, Belmont, Massachusetts, served as Ss. Resident male patients were selected according to the following criteria: (a) primary diagnosis of schizophrenia, without organic involvement, mental deficiency, or alcoholic addiction and (b) age range of 17-35. The selected patients were subdivided, on the basis of staff diagnosis, into paranoid and nonparanoid groups, with 24 patients in each. The nonparanoid group included the following diagnostic categories: simple, 2; catatonic, 2; affective, 2; undifferentiated, 18. In addition, each patient received the Progressive Matrices test (Raven, 1949) and a forced choice version of the Wechsler Adult Intelligence Scale (WAIS) Vocabulary test.

The groups were matched so that there were no significant differences between the paranoid and nonparanoid samples with respect to WAIS (Forced Choice) Vocabulary test, Progressive Matrices, socioeconomic status (based upon father's occupation), and education and age at time of testing. There were no significant differences between the two groups for premorbid level of adjustment based on the Elgin Prognostic Scale (Becker, 1955) and the Prognostic Rating Scale (Phillips, 1953).

A comparison group of 24 male college students were drawn from the evening session division of Boston University. The student group differed from the patients in having higher average WAIS verbal intelligence scores, being on the average 4-5 yr. younger, and coming from higher socioeconomic class.

Procedure

The *E* met with each patient on the ward in order to solicit participation in the study. The game was described by providing the patient with a card containing an actual game matrix and then proceeding to describe the choice options, the payoffs, and the structural interdependence for the two players (see Figure 1). After the patient demonstrated understanding of the game, *E* would ask if he would consent to come to the laboratory to participate in a game-playing study.

The procedure for soliciting the college student sample was somewhat different. Students in the class setting were asked to volunteer to participate in a game-playing study. The laboratory was designed to accommodate six Ss at one time. The students were provided with the same experimental instructions and rewards as the patients. A more detailed description of this laboratory setting has been presented elsewhere (Harford, 1965).

Apparatus and Experimental Situation

The laboratory setting consisted of a large rectangular room with two large mobile booths set up at opposite sides of the room. These booths shielded Ss from seeing or talking to each other. Each booth housed a small electrical switch box which was wired to a master panel, permitting *E* to intercept and transmit game choices to *S* without his knowledge.

The S's panel box was equipped with two rows of lights (two lights per row) and two switches located directly below the lights.

The payoff matrix of the game appears in Figure 1. Player 1 (*S*) chose between red and green (rows) while Player 2 (the stooge) chose between red and green (columns). Depending on the combination of choices, the players either won or lost chips. Player 1's wins and losses were circled. At the end of the game Ss exchanged the chips they had won for cigarettes or candy. The structure of the game was such that both players could win (via red choices), or both could lose (via green choices), or one could win and the other lose.

Saint and Sinner Strategies

The Ss were randomly assigned to one of the two experimental conditions. In one condition (lapsed-saint strategy, Harford & Solomon, 1967) the confederate chose cooperatively for the first three trials (unconditional cooperation) and then switched to conditional cooperation for the remainder of the game (the program cooperated only if *S* cooperated on the immediately preceding trial). On the other hand, in the second condition (reformed-sinner strategy) the confederate chose noncooperatively for the first three trials, then switched to unconditional cooperation for the next three trials, and finally changed to conditional cooperation for the remainder of the game. Both strategies are similar, the only difference being that the reformed sinner initiates the game with three noncooperative choices and thereafter repeats the strategy of the lapsed saint.

The major dependent variables in the present study were trust, risk taking, exploitation, and suspicion, as derived and defined from an analysis of *S*'s expectations of the other's behavior, together with *S*'s own behavior in the game (Solomon, 1960). Trust was defined as a cooperative choice based on an expectation of cooperation from the other player, whereas suspicion was a noncooperative choice based on an expectation of noncooperation. Exploitation was defined by a noncooperative choice based on an expectation that the other player would cooperate, whereas risk taking was a cooperative choice based on an expectation of noncooperation.

RESULTS

Separate analyses were performed on each of the following measures: cooperation, trust, suspicion, risk taking, and exploitation. These measures were not completely independent of each other. For example, trust and exploitation have in common an expectation of cooperation, whereas trust and risk taking have in common a cooperative choice. Table 1 provides a set of intercorrelations of the game-playing measures based on the patient sample. The intercorrelations among trust, suspicion, and

TABLE 1

INTERCORRELATIONS BETWEEN GAME INDEXES FOR THE SAMPLE OF SCHIZOPHRENIC PATIENTS

Variable	T	Risk taking	Suspicion	Exploitation	Cooperative choice
Trust	1.00				
Risk taking	-.33*	1.00			
Suspicion	-.61**	-.30*	1.00		
Exploitation	-.44**	.13	-.25	1.00	
Cooperative choice	.85**	.20	-.80**	-.38**	1.00
Cooperative expectation	.80**	-.28	-.83**	.18**	.68**

Note.—*N* = 48 patients.

* $p < .05$.

** $p < .01$.

cooperation were high, whereas those for exploitation and risk taking were low.

Trial 1

A separate analysis was made for the first trial choice for the three experimental groups. Trial 1 provides a relatively pure measure of *S*'s readiness to trust an unknown other. There were no significant differences between the three groups for any of the measures. Taking the sample as a whole the breakdown was as follows: trust 37%, risk taking 10%, suspicion 29%, and exploitation 24%. Fifty percent of the *Ss* chose to cooperate on the first trial.

Effect of Strategy Change: Trials 1-4 versus 4-8

Experiments designed to test the effects of variations in the confederate's strategy on *S* indicate that *S*'s strategy change may be a more sensitive indicator of treatment effects than the mean number of a particular type of response (Swingle, 1967). To assess *S*'s reactions to the confederate's abrupt change (at Trial 4), analyses of covariance were applied to Trials 5-8.⁵ The number of choices made during Trials 1-4 served as covariates and the number of choices made during Trials 5-8 served as criterion (Winer, 1962). The game measures were transformed by the common logarithm.

Cooperation. The analysis of covariance yielded a significant Groups \times Strategies interaction ($F = 4.48$, $df = 2/65$, $p < .05$) for

⁵ Although both the saint and the sinner abruptly change their strategy on Trial 3, *S* (who chooses simultaneously) made his trial choice without any knowledge of the confederate's change. Thus Trials 1 through 4 were treated as a common phase.

cooperation on Trials 5-8. Table 2 presents the adjusted means for Trials 5-8, together with the unadjusted means for Trials 1-4 and 5-8. Multiple tests between the means were based on the average effective error (Winer, 1962). The results indicated that the students were more reactive to the abrupt change in the confederate's behavior than either of the patient groups. The students were less cooperative with the lapsed saint than with the reformed sinner ($p < .01$). The paranoid schizophrenics were significantly more cooperative with the lapsed saint than either the nonparanoid schizophrenics ($p < .10$) or the students ($p < .05$). Although there were no significant changes within each patient group with respect to the strategies, it is of interest to note that the differential impact of the strategies on cooperative behavior was more similar for the nonparanoid schizophrenics and the students than for the paranoid schizophrenics.

Trust. The analysis of covariance yielded a significant Groups \times Strategies interaction ($F = 3.72$, $df = 2/65$, $p < .05$) for trust on Trials 5-8. The adjusted means for Trials

TABLE 2
ADJUSTED MEANS ON COOPERATION FOR TRIALS 5-8
AND UNADJUSTED MEANS FOR
TRIALS 1-4 AND 5-8

Mean	Paranoids		Nonparanoids		Students	
	LS	RS	LS	RS	LS	RS
Unadjusted Trials 1-4	.38	.44	.51	.36	.39	.27
Unadjusted Trials 5-8	.49	.43	.40	.42	.24	.46
Adjusted Trials 5-8	.50	.39	.32	.44	.25	.54

Note.—LS = lapsed saint; RS = reformed sinner.

TABLE 3

ADJUSTED MEANS ON TRUST FOR TRIALS 5-8 AND UNADJUSTED MEANS FOR TRIALS 1-4 AND 5-8

Mean	LS	RS	LS	RS	LS	RS
Unadjusted						
Trials 1-4	.31	.41	.34	.32	.27	.22
Trials 5-8	.43	.37	.37	.35	.24	.45
Adjusted						
Trials 5-8	.43	.31	.35	.34	.27	.51

Note.—LS = lapsed saint; RS = reformed sinner.

5-8, together with the unadjusted means for Trials 1-4 and 5-8, are presented in Table 3. The students were less trusting with the lapsed saint than with the reformed sinner ($p < .05$). The paranoids were less trusting than the students with the reformed sinner ($p < .10$). There were no significant differences within each patient group with respect to the strategies.

Exploitation. The analysis of covariance yielded a significant Groups \times Strategies interaction ($F = 4.07$, $df = 2/65$, $p < .05$) for exploitation on Trials 5-8. Table 4 presents the adjusted means for Trials 5-8, together with the unadjusted means for Trials 1-4 and 5-8. The nonparanoid schizophrenics were more exploitative with the lapsed saint than with the reformed sinner ($p < .01$). The nonparanoid schizophrenics were more exploitative with the lapsed saint than either the paranoid schizophrenics ($p < .05$) or the students ($p < .10$). Although not significant, the patternings of exploitation under both strategies were more similar for nonparanoids and students than for paranoids.

Suspicion and risk taking. The results of the covariance for the measures suspicion and

TABLE 4

ADJUSTED MEANS ON EXPLOITATION FOR TRIALS 5-8 AND UNADJUSTED MEANS FOR TRIALS 1-4 AND 5-8

Mean	Paranoids		Nonparanoids		Students	
	LS	RS	LS	RS	LS	RS
Unadjusted						
Trials 1-4	.25	.12	.21	.19	.18	.19
Trials 5-8	.16	.17	.28	.11	.16	.10
Adjusted						
Trials 5-8	.14	.19	.27	.11	.16	.09

Note.—LS = lapsed saint; RS = reformed sinner.

risk taking did not yield any significant differences.

Trials 9-30

In order to examine the subsequent impact of the strategies the game measures were summed over Trials 9-30.

Cooperation. The analysis of variance yielded a significant main effect for groups ($F = 4.29$, $df = 1/66$, $p < .05$) for cooperation on Trials 9-30. The data indicated that the patients were more cooperative ($M = .98$) than the students ($M = .77$).

Trust. The analysis of variance yielded a significant main effect for groups ($F = 4.23$, $df = 1/66$, $p < .05$) for trust on Trials 9-30. An examination of the means indicated that the patients were more trusting ($M = .89$) than the students ($M = .66$). The interaction Patients versus Students \times Strategies approached significance ($p < .10$). The patients displayed higher levels of trust with the lapsed saint ($M = .99$) than with the reformed sinner ($M = .79$). The students, on the other hand, displayed higher levels of trust with the reformed sinner ($M = .78$) than with the lapsed saint ($M = .55$). An examination of the means for each patient group indicated a greater differential response to the lapsed saint and reformed sinner for the nonparanoid schizophrenics ($M = 1.04$ versus .71, respectively) than was true of the paranoid schizophrenics ($M = .93$ versus .87, respectively).

Exploitation. The analysis of variance of the number of exploitative choices yielded a main effect for patients which approached significance ($F = 3.0$, $df = 1/66$, $p < .10$). The nonparanoid schizophrenics tended to be higher in exploitation ($M = .55$) than the paranoid schizophrenics ($M = .36$).

DISCUSSION

Paranoid versus nonparanoid differences. The two samples did not differ with regard to their game behavior on the first trial. An analysis of the effects of the abrupt change in the confederate's strategy revealed differences between the two patient groups. The lapsed saint produced higher levels of cooperation in the paranoid sample than in the nonparanoid

sample. The reformed sinner produced lower levels of exploitation in the paranoid sample when compared with the nonparanoid sample. These initial reactions to the two strategies tended to be less salient during the remainder of the game (Trials 9-30). An apparent explanation stems from the merging of the two strategies into a similar strategy of conditional cooperation. Despite this, however, there were some subtle differences between the patient samples. The nonparanoids tended to be higher in exploitation than the paranoids, and were more responsive to the differences between the two strategies.

Perhaps the major issue at stake for the paranoid schizophrenics in the present experimental situation is not when and whom to trust, but rather a concern with the balance of power. Ovesey (1955) has proposed that internalized conflicts with regard to achieving interpersonal power are a key feature of the paranoid orientation. The paranoids' strategy is designed to test and redefine the existing balance of power. Thus, when the sinner is punitive and tough, the paranoid is docile and cooperative; with the sinner's abrupt shift to a soft unconditional benevolence, the paranoids react by shifting in the direction of an exploitative and dominating strategy. With the saint the reverse is true. The paranoids are initially exploitative, but when the saint shifts to a strategy of punishing noncooperation, the paranoids rapidly shift to a cooperative pattern.

The game-playing strategies of the paranoids also fit the clinical description of the authoritarian personality (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950). Adorno et al. described the authoritarian as a person who idealized strength and toughness but has contempt for weakness and tenderness. They imply that the authoritarian is likely to be exploitative and domineering when he has greater power (or is not resisted), but submissive and ingratiating when in a subordinate relationship. The conception of the paranoid person as one who is conflicted with respect to power relationships has applicability to normal as well as schizophrenic samples.

One key implication of these findings is

that the paranoid schizophrenic's interpersonal stance is not indiscriminate and can be modified by variations in the strategy of the other party. In order to illuminate other clinical aspects of suspicion in paranoia, future research designs could systematically vary such interpersonal dimensions as the status, sexual identity, and perceived similarity of the other party.

Patient versus student differences. The student sample was included in order to provide a comparative point of reference rather than as a control. The students represent a limiting case in the point of comparisons: Are the prisoner's dilemma game and strategy manipulations sensitive to the differences in these two sample populations? The findings indicated three major differences between the combined patient sample and the students: (a) the students showed a higher level of reactivity to the confederate's strategy shift than either of the patient samples; (b) the students showed higher levels of noncooperation than either of the patient samples; and (c) the reformed sinner tended to be more effective in establishing a trusting relationship with the students than with the patients. These findings contrast with the study by Knapp and Podell (1968) who did not obtain differences between psychotic patients and college students. The differences obtained in the present experiment were largely a function of (a) a more well-defined psychotic population and (b) type of strategy. Knapp and Podell (1968) pointed out that variations in cooperative strategies are more likely to reveal personality differences than the strictly competitive strategies.

The finding that the patients as a group were more cooperative and trusting may reflect the fact that the students adopted a more competitive set in the game. The students' higher levels of reactivity to the strategy shift of a confederate suggests that they were more sensitive to interpersonal change in others. Moreover, they were selectively sensitive to a shift from a competitive to a cooperative strategy. Whether the differences between the student and patient samples can be attributed to the nature of the schizophrenic illness, or to other possible confounding variables such

as the effects of institutionalization (Goffman, 1961; Silverman, 1965), social stigma (Goffman, 1963), or the social psychology of the experimental situation (Rosenthal, 1964) can not be determined.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- ARIETI, S. Introductory notes on the psychoanalytic therapy of schizophrenics. In A. Burton (Ed.), *Psychotherapy of the psychoses*. New York: Basic Books, 1961.
- BECKER, W. C. The relations of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctoral dissertation, Stanford University, 1955.
- BIXENSTINE, V. E., & WILSON, K. V. Effects of level of cooperative choice by the other player in choices in a Prisoner's Dilemma Game. Part II. *Journal of Abnormal and Social Psychology*, 1963, 67, 139-147.
- DEUTSCH, M. Trust and suspicion. *Journal of Conflict Resolution*, 1958, 4, 265-279.
- DEUTSCH, M. Cooperation and trust: Some theoretical notes. *Nebraska Symposium on Motivation*, 1962, 10, 275-320.
- ERIKSON, E. H. Growth and crises of the healthy personality. In C. Kluckhohn & H. Murray (Eds.), *Personality in nature, society and culture*. (2nd ed.) New York: Knopf, 1953.
- FROMM-REICHMANN, F. Notes on the development of treatment of schizophrenics by psychoanalytic psychotherapy. *Psychiatry*, 1948, 11, 263-273.
- GOFFMAN, E. *Asylums*. New York: Doubleday, 1961.
- GOFFMAN, E. *Stigma: Notes on the management of spoiled identity*. New Jersey: Prentice-Hall, 1963.
- HARFORD, T. Game strategies and interpersonal trust in schizophrenics and normals. Unpublished doctoral dissertation, Boston University, 1965.
- HARFORD, T., & SOLOMON, L. "Reformed sinner" and "lapsed saint" strategies in the Prisoner's Dilemma game. *Journal of Conflict Resolution*, 1967, 11, 104-109.
- KATZ, M. M., COLE, J. O., & LOWERY, H. A. Non-specificity of diagnosis of paranoid schizophrenia. *Archives of General Psychiatry*, 1964, 11, 197-202.
- KLEIN, E. B., & SOLOMON, L. Agreement response tendency and behavioral submission in schizophrenia. *Psychological Reports*, 1966, 18, 499-509.
- KNAFF, W. M., & PODELL, J. E. Mental patient, prisoner and student with simulated partners in a mixed-matrix game. *Journal of Conflict Resolution*, 1968, 2, 235-241.
- LORR, M. Factors descriptive of psychopathology and behavior of psychotics. *Journal of Abnormal and Social Psychology*, 1955, 50, 78-86.
- LORR, M., O'CONNOR, J. P., & STAFFORD, J. W. The psychotic reaction profile. *Journal of Clinical Psychology*, 1960, 16, 241-245.
- LORR, M. *Explorations in typing psychotics*. London: Pergamon Press, 1966.
- LYNN, R. Russian research in schizophrenia. *Psychological Bulletin*, 1963, 60, 486-498.
- MCGONAGHY, N. Modes of abstract thinking and psychosis. *American Journal of Psychiatry*, 1960, 117, 106-110.
- OVSEY, L. Pseudohomosexuality, the paranoid mechanism and paranoia. *Psychiatry*, 1955, 18, 163-173.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease*, 1953, 117, 515-525.
- RAVEN, G. C. *Coloured Progressive Matrices*. (Orig. publ. 1947) London: Lewis, 1949.
- ROSENTHAL, R. The effect of the experimenter on the results of psychological research. In B. Maher (Ed.), *Progress in experimental personal research*. New York: Academic Press, 1964.
- SCODEL, A. Induced collaboration in some non-zero-sum games. *Journal of Conflict Resolution*, 1959, 3, 114-119.
- SHAKOW, D. Psychological deficit in schizophrenia. *Behavioral Science*, 1963, 8, 272-305.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 352-379.
- SILVERMAN, J., BERG, P., & KANTOR, R. Some perceptual correlates of institutionalization. *Journal of Nervous and Mental Disease*, 1965, 141, 651-657.
- SOLOMON, L. Influence of some types of power relationships and game strategies upon the development of interpersonal trust. *Journal of Abnormal and Social Psychology*, 1960, 61, 223-230.
- SWINGLE, P. G. Response variability in non-zero-sum games. Paper presented at the meeting of Eastern Psychological Association, Boston, April 1967.
- WIENER, J. *Psychodiagnosis in schizophrenia*. New York: Wiley, 1966.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 8, 1968)

FALSE FEEDBACK AND LONGEVITY OF THE CONDITIONED GSR DURING EXTINCTION:

SOME IMPLICATIONS FOR AVERSION THERAPY¹

KARL P. KOENIG² AND DAVID DEL CASTILLO

University of New Mexico

The effects of two cognitive variables on GSR extinction rates were investigated. Study I replicated earlier findings that Ss made aware of the onset of non-shocked trials exhibited extinction of a conditioned emotional response more rapidly than Ss not made aware. Studies I and II further demonstrated that extinction, even under conditions of awareness, can be significantly retarded when false feedback concerning the S's current emotional responsivity is supplied during extinction trials.

Fundamental to most behavioral treatment of sexual deviance has been the association of aversion with stimuli thought to control sexual behavior (Feldman, 1966; Rachman, 1961). Thorpe and Schmidt (1964), for example, paired noxious shock with photographs of nude men in a case of male homosexuality. There have also been a number of attempts to elaborate the basic classical conditioning paradigm to include avoidance learning (Feldman & MacCulloch, 1965), escape learning (Barker, 1965), and aversion-relief training (Fisher, 1968). Regardless of the several variations, however, successful behavior modification must be presumed to depend upon the establishment and maintenance of a vigorous and negative emotional response to the conditioned sexual stimuli.

The use of aversive conditioning as a therapeutic procedure rests on the assumption that the conditioned emotional response (CER) established in therapy will continue to resist extinction after S has left the treatment setting. In the absence of firm clinical evidence to the contrary, an alternative assumption is instead that the deviant client will learn to respond fearfully to sexual

stimuli when it is appropriate to do so (in the therapy setting) yet continue to respond nonanxiously in those circumstances where fear of aversive consequences would be inappropriate (locales other than the clinic). One would expect humans to discriminate readily between those occasions with zero probability of aversive unconditioned stimulus (US) occurrence and those with a probability greater than zero. If the goal of aversion therapy is to generate specific conditioned fear under the control of particular sexual stimuli, but independent of other significant stimuli such as the presence or absence of shock electrodes, then present techniques would appear inadequate for the task. Learned fear which persists across nonreinforced presentations of the conditioned stimulus (CS) depends on the inaccessibility of precisely that information which enables S to make good probabilistic estimates of the occurrence of punishment. The variables suggested by Feldman (1966) as useful in producing resistance to extinction (e.g., distributed versus massed trials) are likely to have a major effect only if Ss remain ignorant of the onset of non-reinforced trials.

Empirical evidence is available which indicates that rapid, if not immediate, extinction of the CER occurs when Ss are informed that subsequent presentations of the CS are no longer to be followed by shock. The effect has been demonstrated in conditioned heart-rate studies (Chatterjee & Erikson, 1962; Notterman, Schoenfeld, & Bersh, 1952) and in GSR conditioning (Bridger & Mandel,

¹ This research was supported by Research Grant MH-12854-02 and by the Research Allocations Committee of the University of New Mexico. Grateful acknowledgment is owed to Thomas Vernardos who assisted in the experiment, to the Research Computer Center for assistance in statistical analysis, and to Stanley Hipwood who drew the figures.

² Requests for reprints should be sent to Karl P. Koenig, Department of Psychology, University of New Mexico, Albuquerque, New Mexico.

1965; Cook & Harris, 1937; Mowrer, 1938; Silverman, 1960; Wickens, Allen, & Hill, 1963). Inasmuch as clinical patients are also "informed" (by implication) of when aversive events are probable and when not, investigators of therapeutic aversive conditioning should expect little resistance to extinction beyond the confines of the clinic.

In view of the relationship between laboratory studies of CER extinction and current therapeutic practices, the present investigation was designed to determine whether a CER could be maintained during extinction despite instructions which informed Ss of the permanent termination of shock following acquisition trials. The Ss also viewed, by means of a meter, experimentally manipulated false feedback represented as their emotional responses to CS trials. Thus, each S was conceived to be serving as his own response-cueing model. Since it has been shown that emotional responses can be both acquired and extinguished through exposure to models (Barnett & Benedetti, 1960; Berger, 1962), it was predicted that Ss would *continue* to respond fearfully to a stimulus if supplied with evidence that they were *already* responding in a fearful manner.

STUDY I

Method

Subjects. The initial sample consisted of 97 undergraduates. Failure to meet either of two criteria resulted in the elimination of 17 Ss. Five did not meet the criterion of scorable records due either to apparatus failure or to erratic resistance changes which masked their responses. The remaining 12 were eliminated for failure to respond above .1 log change in conductance to the test stimulus prior to extinction trials. The final sample consisted of 32 women and 48 men assigned randomly to the eight experimental subgroups.

Apparatus. A Lafayette dermohmeter (7601A) was used in conjunction with a Harvard Electronic Recording Module (350) to produce a continuous ink recording of resistance change. A second pen recorded time and stimulus events. The equipment, including timers, relays, and shock stimulator, was located in a control room connected by a two-way mirror to the S room.

Nickel-plated finger electrodes were taped to S's first and third right hand fingers and connected to the dermohmeter. His right arm was secured by an elastic bandage to the arm of a recliner chair to reduce GSR artifacts. Shock was delivered to the

left hand in .5-sec. impulses in a range of .0-5.5 ma. through a 100 K Ω resistor in series with S.

The S sat in a small but comfortable room. A headset and intercom permitted two-way communication with E. The chair faced a panel of three jewel lights, 4.5 ft. away at eye level, and visible through the two-way mirror when illuminated. Lights were timed for 2-sec. durations. The reinforced light (positive conditioned stimulus, CS+) was orange, while the nonreinforced stimuli (negative conditioned stimuli, CSs-) consisted of two white lights equal in size and brightness to the CS+. The CS+ was easily discriminable from the CSs-. Also visible, when illuminated by E during extinction trials, was a 4 x 3 in. black-on-white meter set 4 in. below the lights. Each needle movement of the meter started at zero and peaked halfway across the dial before returning slowly to zero.

Procedure. On arrival, S was seated, headphones and electrodes were attached, and instructions to remain very still were given. After apparatus adjustments were made the room lights were dimmed to enhance attention to the stimuli. The experiment consisted of the following four phases:

1. **Adaptation.** Ten positive conditioned stimuli (CSs+) and 20 CSs- were presented in a fixed-random order with a 25-sec. interval between presentations.

2. **Shock determination.** Each S was asked to assist in finding a barely tolerable level of shock. The individually determined level was then used throughout acquisition (obtained range = .5 to 5 ma.).

3. **Acquisition.** Ten CSs+ and 20 CSs- were presented in a new fixed-random order with the same intertrial interval. Each CS+ was followed immediately by a shock of .5-sec. duration 100% of the time. After all 30 stimuli were presented, an eleventh CS+ without shock was presented to determine conditioning (test trial).

4. **Extinction.** Twelve CSs+ and 24 CSs- were presented in a new order with the same interval. No shock was delivered at any time during extinction. No differential conditions were introduced until the extinction phase.

Information conditions. Following the test trial, but prior to extinction trials, one-half of the Ss (informed group) received instructions designed to assure them that the remaining stimulus presentations would not involve shock. Abstracted, instructions for the informed condition were as follows. "I am coming in now to remove the shock electrodes. From this point on, none of the lights will be followed by shock." The E joined S and removed the electrodes.

The remaining Ss received no indication that shock would be terminated. To control for E interruption, however, noninformed Ss were told that an adjustment was to be made in the amplifying system. The E then entered the room and appeared to make such an adjustment.

Meter conditions. The presence or absence of meter activity visible to Ss and different sets of instructions resulted in four variants of meter conditions.

These were balanced between the informed and non-informed conditions. Meter instructions were given following information instructions but prior to extinction trials. Meter activity began immediately upon termination of each CS+ and was completed within 7 sec. after its inception. Meter activity followed only the first six extinction trial CSs+ and during this time the meter light was on constantly for all meter conditions which were the following:

1. *Meter-feedback groups.* In addition to exposure to meter activity, instructions were given which are reproduced in part here.

during this part of the experiment you will be able to see your own galvanic skin response to each light as it is presented. . . . If the needle rises above 10 units [one-eighth the possible distance] this will indicate you have given a fear or anxiety response to the light.

The meter needle in all cases, of course, moved well beyond the mark signifying anxiety to S since it always peaked at approximately one-half the distance between 0 and 120.

2. *Meter-task groups.* Although the meter was manipulated exactly as in the above condition, the instructions differed: "you should pay attention to the meter as well as the lights because I will later ask you some questions about the activity of the meter needle."

3. *Meter-only groups.* No instructions were given concerning the meter but the meter was operated as in Conditions 1 and 2.

4. *No-meter groups.* Neither meter activity nor meter instructions were introduced to these groups.

Groups under Conditions 1, 2, and 3 saw the same meter and, within limits of experimental error, the same meter activity. Only the instructions describing the purpose of the meter differed. The meter did not appear again after the last pairing of the meter with the sixth CS+ during extinction. No explanation for meter termination was offered. The purpose of terminating the meter after the sixth CS trial was to examine the transference of its effects to the last six CS trials.

Results

The GSR unit changes were converted from changes in resistance to log changes in conductance (log CCs) in a manner consistent with that suggested by Haggard (1949), Lacey and Siegel (1949), and Wickens et al. (1963). The raw measure consisted of the maximum resistance change (ohms) within an 8-sec. period after CS+ termination. Positive log CCs reflect decrements in skin resistance. A constant of one was added to each change value before the logarithmic transformation was made. The CS+ extinction trials were grouped in six blocks of

TABLE 1
STUDY I ANALYSIS OF VARIANCE OF LOGARITHMIC CHANGE IN CONDUCTANCE BY INFORMATION AND METER CONDITIONS AND BLOCKS

Source	df	MS	F
Information (I)	1	6.2794	17.946**
Meter (M)	3	1.3022	3.722*
I \times M	3	.0534	
Error	72	.3499	
Blocks (B)	5	1.0710	13.696**
B \times I	5	.3074	3.930**
B \times M	15	.0054	
B \times I \times M	15	.1522	1.946*
Error	360	.0782	

* $p < .025$.

** $p < .005$.

two trials each for analysis and graphical presentation.

An analysis of variance with repeated measures was performed on the log CCs. The summary of the analysis is shown in Table 1 and the associated functions are given in Figure 1. In addition, log CCs associated with the test trial alone were analyzed in order to test the assumption that no real differences existed among the experimental subgroups prior to

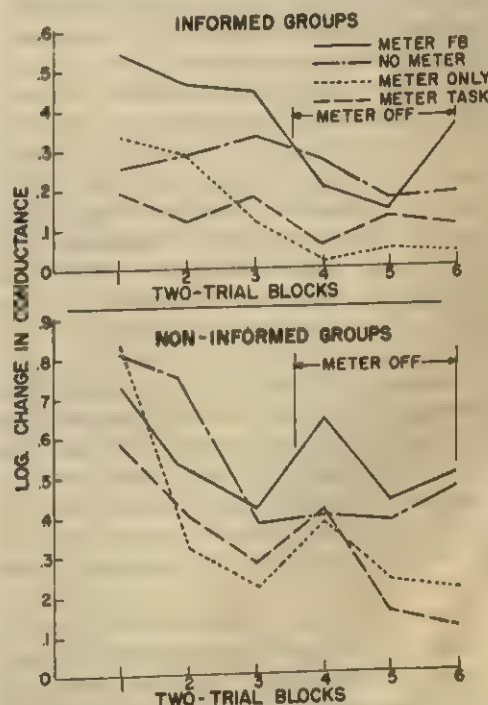


FIG. 1. Study I extinction curves ($N = 10$ in each group).

the introduction of experimental variables. Since no comparisons yielded an F greater than 1, the assumption of initial group comparability was supported. A correlation of .51 ($df = 78$, $p < .01$) was obtained between test-trial responses to the CS and averaged responses to the extinction CSs. Therefore, it was of importance that no significant group differences existed initially.

The significant effects for information and for Information \times Blocks (see Table 1) supported the prediction based on earlier studies that Ss made aware of the termination of shock would extinguish more rapidly and to a lower level than Ss not so informed (see Figure 1). The log CC block means associated with the main effect were noninformed = .45 and informed = .22.

Although meter conditions resulted in a significant overall effect, it was not possible to interpret this finding without examining the individual group means. Duncan's multiple-range test (Edwards, 1960) was employed for this purpose. At the .05 level of significance, and within the informed-of-extinction condition, both the meter-task and the meter-only groups differed from the meter-feedback group in the expected direction. The no-meter group differed from the meter-feedback group in the predicted direction but only between the .10 and .05 levels of significance. The top drawing in Figure 1 shows that the meter-feedback group under the informed condition produced stronger conditioned responses than the no-meter group during the first three blocks of trials, but dropped steeply when the supply of false feedback was terminated prior to the last three blocks of trials.

Tests of individual group means within the noninformed conditions yielded effects similar to those within the informed condition: the meter-feedback group extinction rate was significantly retarded when compared with the meter-only and meter-task groups, but did not differ significantly from the no-meter condition.

Support for the prediction concerning information, and partial support for the expected effects of meter feedback, stimulated the second study. The experiment reported in the following paragraphs was undertaken in order to test the prediction that meter feed-

back will result in slower extinction than a no-meter condition under a constant-informed-of-shock-termination condition, over extended acquisition and feedback trials.

STUDY II

Method

Subjects. The initial sample consisted of 26 undergraduates enrolled in introductory psychology. Failure to meet the criterion of scorable records led to the elimination of 3 Ss. Another 3 Ss were eliminated for failure to respond above .1 log CC to the test trial CS prior to extinction. The remaining 20 Ss consisted of 11 men and 9 women assigned randomly to the two experimental conditions.

Apparatus. The apparatus was identical to that in Study I.

Procedure. Adaptation, acquisition, and extinction trials were conducted, as in Study I, but repeated on 2 consecutive days for each S. On both days, prior to extinction trials, all Ss were informed that shock would no longer occur and the shock electrodes were removed. The no-meter group was not exposed to the meter on either day. The meter-feedback group, however, was shown the meter (with instructions identical to those used in Study I) following all 12 CS+ presentations during extinction on Day 1, and following each of the first 6 CS+ presentations during extinction on Day 2. Thus, the following sequence was observed for all Ss. Day 1 consisted of 10 CS+ and 20 CS- trials in fixed-random order during adaptation; 10 CS+ and 20 CS- trials in another fixed-random order during acquisition; and 12 CS+ and 24 CS- trials in still another order during extinction. Day 2 followed the same pattern except that new fixed-random arrangements of the stimuli were generated.

Results

Skin resistance changes to the CS were computed and transformed into log CCs as described in Study I. Separate t tests were performed on test-trial log CCs between the no-meter and meter-feedback conditions for Day 1 and Day 2. On neither day was there a significant difference between the test-trial means. The absence of a test-trial difference between the two conditions on Day 1, prior to the introduction of experimental treatments, supported the assumption that the two S samples had been conditioned to equivalent levels.

Analysis of variance with repeated measures was performed separately for Day 1 and Day 2 data. The obtained Day 1 F of 4.13 ($df = 1/18$) for the main effect was not statistically significant. The analysis of Day 2 data, however, yielded an F of 7.55 ($df =$

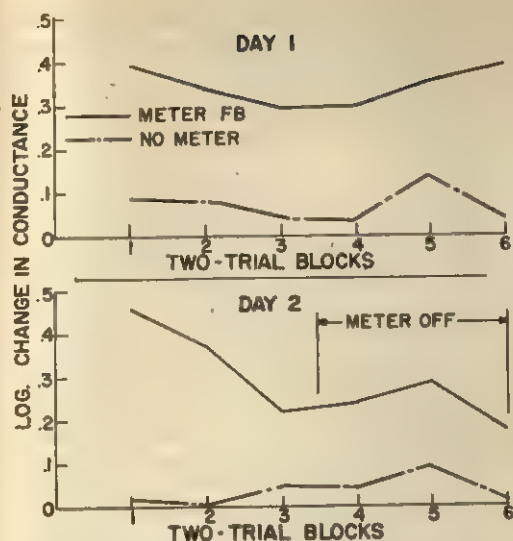


FIG. 2. Study II extinction curves ($N = 10$ in each group).

1/18) which was acceptable beyond the .025 level of significance. In neither analysis did the blocks or Blocks \times Meter effects reach statistically significant levels. The results, compared with the no-meter and meter-feedback means under the informed condition in Study I, supported the prediction that false GSR feedback would effectively retard extinction when the number of feedback trials during extinction was increased.

The data are represented in Figure 2. Unlike meter feedback under the informed condition in Study I, the responses of comparable meter-feedback Ss in Study II did not show rapid decline in responsivity after the first three blocks of extinction trials. This was probably due to the fact that false meter feedback was continued across all extinction trials on Day 1. A relatively steep decline was noted on Day 2 over the first three blocks which leveled off following discontinuation of feedback.

DISCUSSION

Differential rates of extinction were obtained between informed and noninformed groups in Study I. The effect was in the direction predicted and was in general agreement with other experimental reports (e.g., Bridger & Mandel, 1965). Earlier in the paper

a parallel was drawn between experimental Ss aware of shock termination and clinic patients who are also aware of when shock contingencies are likely to operate and when they are not. The finding attributed to information conditions in the present study, and the parallel described above, raises a serious question concerning the probable effectiveness of aversive therapy in maintaining a strong CER which is also resistant to extinction outside the therapeutic setting. Nor is it immediately apparent why instrumentally conditioned avoidance responses (vigorously proposed by Feldman, 1966) would be less susceptible to rapid extinction in view of the average clinic S's ability to discriminate presence and absence of avoidance contingencies. The omission of systematic consideration of cognitive variables such as awareness seems to strengthen the recent criticism of conditioning therapies made by Breger and McGaugh (1965).

False meter feedback, labeled explicitly for S as his own galvanic skin response to the previously shocked light, was shown to be an effective procedure for retarding extinction rates even in Ss aware of the onset of nonshocked trials. It will be recalled that differences between the false-feedback and no-meter conditions reached significance in the second study only with the addition of extended acquisition and extinction trials. The requirement of additional trials to achieve a differential effect does not, however, seriously weaken the demonstration. Therapy by aversive conditioning has generally consisted of a series of alternated acquisition and extinction trials. Each time the patient leaves the office he is, in effect, entering extinction trials if either he imagines or physically encounters those stimuli to which aversion has recently been conditioned. An examination of the extinction data in Study II indicates that repeated exposure to nonshocked trials, even when interspersed with reinforced trials, led to successively diminished conditioned responses in the absence of feedback. Conversely, there appeared to be a lesser decline in the magnitude of the response during extinction in meter-feedback Ss and an increase in response stability. Further research will be required in order to determine the feedback

format and values and the frequency of acquisition trials which tend to maximize resistance to extinction.

REFERENCES

- BARKER, J. C. Behavior therapy for transvestism: A comparison of pharmacological and electrical aversion techniques. *British Journal of Psychiatry*, 1965, 111, 268-276.
- BARNETT, P. E., & BENEDETTI, D. T. A study in "vicarious conditioning." Paper presented at the meeting of the Rocky Mountain Psychological Association, Glenwood Springs, Colorado, May 1960.
- BERGER, S. M. Conditioning through vicarious instigation. *Psychological Review*, 1962, 69, 450-466.
- BREGER, L., & MCGAUGH, J. L. Critique and reformulation of "learning theory" approaches to psychotherapy and neurosis. *Psychological Bulletin*, 1965, 63, 338-358.
- BRIDGER, W. H., & MANDEL, I. J. Abolition of the PRE by instructions in GSR conditioning. *Journal of Experimental Psychology*, 1965, 69, 476-482.
- CHATTERJEE, B. B., & ERICKSEN, C. W. Cognitive factors in heart-rate conditioning. *Journal of Experimental Psychology*, 1962, 64, 272-279.
- COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology*, 1937, 21, 202-210.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.
- FELDMAN, M. P. Aversion therapy for sexual deviations: A critical review. *Psychological Bulletin*, 1966, 65, 65-79.
- FELDMAN, M. P., & MACCULLOCH, M. J. The application of anticipatory avoidance learning to the treatment of homosexuality. I. Theory, technique, and preliminary results. *Behaviour Research and Therapy*, 1965, 2, 165-184.
- FISHER, D. Clinical applications of aversive conditioning. Paper presented at the meeting of the Western Psychological Association, San Diego, March 1968.
- HAGGARD, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *Journal of Experimental Psychology*, 1949, 39, 378-392.
- LACEY, O. L., & SIEGEL, P. S. An analysis of the unit of measurement of the galvanic skin response. *Journal of Experimental Psychology*, 1949, 39, 122-127.
- MOWSER, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychological Review*, 1938, 45, 62-91.
- NOTTERMAN, J. M., SCHOENFELD, W. W., & BERSH, P. J. Conditioned heart rate response in human beings during experimental anxiety. *Journal of Comparative and Physiological Psychology*, 1952, 45, 1-8.
- RACHMAN, S. Sexual disorders and behavior therapy. *American Journal of Psychiatry*, 1961, 118, 235-240.
- SILVERMAN, R. E. Eliminating a conditioned GSR by reduction of experimental anxiety. *Journal of Experimental Psychology*, 1960, 59, 122-125.
- THORPE, J. G., & SCHMIDT, E. Therapeutic failure in a case of aversion therapy. *Behaviour Research and Therapy*, 1964, 1, 293-296.
- WICKENS, D. D., ALLEN, C. K., & HILL, F. A. Effects of instructions and UCS strength on extinction of the conditioned GSR. *Journal of Experimental Psychology*, 1963, 66, 235-240.

(Received July 8, 1968)

RECALL AND RECOGNITION FREE LEARNING IN SCHIZOPHRENICS¹

GILEAD NACHMANI AND BERTRAM D. COHEN²

Rutgers—The State University

Recall and recognition learning tasks were administered to 30 schizophrenic and 30 nonschizophrenic psychiatric inpatients. The schizophrenics were inferior to the nonschizophrenics in the number of correct items reproduced on the recall tests ($p < .05$), but the two groups did not differ significantly in the number of correct items identified on the recognition tests. The schizophrenics made more intrusion errors on both the tests of recall ($p < .01$) and recognition ($p < .05$). Compared to the nonschizophrenics, the schizophrenics' intrusion errors were more marked on the recall than on the recognition tests: the Groups \times Tasks interaction was significant ($p < .01$). The results were related to a theory of overinclusion in schizophrenia, previously developed to account for schizophrenic deficits in a communication task.

Cameron (1938) described cognitive performance in schizophrenia as overinclusive. This is the notion that schizophrenics have difficulty maintaining the conceptual boundaries of task situations; hence, responses inappropriate to the task at hand are likely to intrude into the schizophrenic's thinking and communication. As a consequence, thought and speech become less precise and coherent, more vague and fragmentary.

There have been a number of attempts to explain the phenomena of overinclusion in schizophrenia by relating them to faulty information processing mechanisms (Lang & Buss, 1965; Yates, 1966). Such models typically construe schizophrenic patients as suffering from some defect in a hypothetical "screening" mechanism. Normally, this mechanism functions to filter-out "noise" from internal sources (irrelevant associations), or from external sources (irrelevant stimuli). Some specific models are those proposed by Payne, Matussek, and George (1959), McGhie and Chapman (1961), and Shakow (1962).

A more recent model, also including the

notion of a filter mechanism, is one proposed by Rosenberg and Cohen (1966) and applied to schizophrenic performance in a communication task by Cohen and Camhi (1967). The Cohen and Camhi study was designed to compare the performance of schizophrenic and normal Ss in "speaker" and in "listener" roles. In the communication task, the speaker's job was to provide clue words that would make it possible for his listener to distinguish referent from nonreferent stimuli. The listener's job was to choose the referent stimulus from each two-stimulus array, basing his choices on the speaker's clue words. Using as a measure the proportion of correct listener's choices of the speaker's referents, the results indicated that schizophrenic speakers were inferior to normals, but schizophrenic listeners approached the accuracy of normals. The communication deficit, therefore, appeared to rest mainly in the role of the speaker.

Speaker and Listener Models

What is there about the particular requirements of the speaker's task such that a schizophrenic deficit was found in the speaker role but not in the listener role? Rosenberg and Cohen (1966) proposed that the speaker role involves a two-stage process, while the listener role involves a one-stage process. According to this theory, the speaker's role begins with a *sampling stage* in which a response from the speaker's repertoire is selected randomly from

¹ This research was supported by National Science Foundation Grant No. GS 723 for research on verbal behavior as interpersonal communication. The authors are grateful to Seymour Rosenberg for many useful suggestions in the development of this paper.

² Requests for reprints should be sent to Bertram D. Cohen, Department of Psychiatry, Rutgers University Medical School, New Brunswick, New Jersey 08903.

the distribution of possible clue words associated with the referent stimulus alone. In the *comparison stage*, which always follows the sampling stage, the relative associative strengths linking each of the stimulus objects (referent and nonreferent) to the sampled response determines the probability that the speaker will decide to emit or reject the sampled response. If, as a result of the comparison stage, the sampled response is rejected as an inadequate clue word, the speaker again samples a clue word and then repeats the comparison process. This hypothetical sequence (sampling stage \rightarrow comparison stage) continues to recycle until a satisfactory clue word is decided on by the speaker, and he emits this word as a response. It is in the comparison stage that the speaker must reject ambiguous and misinformative associations. What is implicit in the concept of the comparison stage is its function as a filter mechanism. The speaker must reject clue words that will not have, for the listener, discriminably higher associative strength to the referent than to the nonreferent. The listener's choice of the speaker's referent was also assumed to be governed by the relative associative strengths of the speaker's clue word to each of the two stimulus objects (the referent and the nonreferent).

It will be recalled that in the Cohen and Camhi study, the schizophrenics were found to be defective as speakers, but approached normal accuracy as listeners. When, according to the theory, sampling and comparison stages were *both* involved (speaker process), schizophrenic performance suffered. However, when only the comparison stage was thought to be operative (listener process), the schizophrenics showed no significant inadequacy. Cohen and Camhi (1967) speculated that the specific weakness of the schizophrenic was in his ability to make comparison stage decisions in the speaker role, but not in the listener role. They explained this apparent contradiction in the following way:

Faulty comparison in the speaker role . . . may reflect deficiencies in the ability of schizophrenics to integrate the comparison with the sampling stage of the two-stage speaker process. Thus, comparison combined with sampling (speaker comparison) may be defective, even though comparison alone (listener comparison) remains relatively unimpaired [p. 245].

Recall and Recognition

The speaker and listener tasks are analogous, respectively, to the traditional recall and recognition tests used in studies of rote learning. The essential operational difference between tests of recall and recognition is that in recall *S* is required to reproduce items from a previously presented training list, while in recognition *S* is required only to judge whether items, produced for him by *E*, were or were not on the previously presented training list.

Rosenberg and Cohen (1966) suggested that performance on recall tasks is determined by a two-stage process similar to the speaker model, while performance on recognition tasks is determined by a one-stage process similar to the listener model. In recall, *S* first samples an item from an underlying set that consists of list and extralist subsets. The *S* then goes on to the comparison stage which determines the probability that the sampled item will be emitted as a response. In the comparison stage, *S* decides whether the sampled item was, or was not, in the list subset. Hypothetically, this is done by "comparing the associative strengths of the sampled items with a constant representing a threshold value of associative strength [p. 228]." In recognition, only the comparison stage is involved, as *E* himself performs the sampling for *S*.

One might expect, therefore, that schizophrenics will show deficit on recall tasks but not on recognition, because recall, like the speaker task, presumably involves an integration of both sampling and comparison stages; while recognition, like the listener task, only involves a comparison stage. The schizophrenic's faulty ability to make comparison stage decisions in the two-stage recall task implies that his judgments of an item's list membership will be more unreliable than such judgments by a nonschizophrenic *S*. This can result in (a) overexclusion (Chapman, 1961): frequent erroneous decisions to reject a correct item—leading to lowered correct response scores; and (b) overinclusion: frequent failures to reject incorrect, extralist items—leading to increased intrusion error scores. In contrast, neither of these effects are expected on the one-stage recognition task.

Although no research has been conducted which has directly compared recognition versus recall tests of learning in schizophrenics, Spence, Goodstein, and Lair (1965) observed that schizophrenics performing a recognition task were less distracted by verbal reinforcement than were schizophrenics performing a recall task. Spence et al. suggested that the recognition task demanded less active effort and less sustained concentration, and could, therefore, have made the schizophrenic Ss' performance less vulnerable to any distracting effects the reinforcers may have had.

The purpose of the present study was to provide data comparing schizophrenic and nonschizophrenic Ss on recall and recognition tests of retention in a free learning task.

METHOD

Test Materials

The items for the recall and recognition tasks were obtained in the following manner. Three hundred Rutgers undergraduates, all males, in introductory psychology classes, were asked to list within a 60-sec. period, all of the items they could think of in each of two categories: one was "states" of the United States and the other was "edible plants." This was done to establish frequency norms for two conceptual sets. In all, 50 items (the 50 states) were obtained for the states category, and 54 items for the edible plants category. The percentage of students listing each item was tabulated for the items of both distributions. The distributions were then divided into deciles, and 15 items were selected from each distribution to make up the two training lists used in the study; for each conceptual set, one item was randomly selected from each of the 10 deciles, while the remaining 5 items were randomly selected from the middle three deciles (the fourth through sixth).

Subjects

The Ss were 30 schizophrenic and 30 nonschizophrenic male psychiatric inpatients at the Veterans' Administration Hospital in East Orange, New Jersey. Diagnoses for all of the patients were obtained at psychiatric staff conferences approximately 1 wk. after the patient was admitted to the hospital. The two groups of patients were comparable in age and education (see Table 1) as well as medication. All of the Ss were being treated with comparable dosages of Mellaril (dose to body weight). No S had a history of more than three psychiatric hospitalizations since discharge from military service. All of the schizophrenics had a history of psychotic symptoms since military service. The schizophrenics were non-paranoid, chronic, undifferentiated. The nonschizophrenic group was composed of Ss diagnosed as

TABLE 1

AGE AND EDUCATION OF THE SCHIZOPHRENIC AND NONSCHIZOPHRENIC GROUPS

Group	Age		Education	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Schizophrenic	32.1	4.8	10.1	1.6
Nonschizophrenic	33.9	5.9	11.8	1.9

Note.—*N* = 30 for each group.

anxiety neurotics with no indication of psychotic symptoms.

Procedure

Training. The *S* was told that he would see a list of words, one at a time. As each word appeared, he was instructed to say it aloud and try to remember it. Each word was presented for 2 sec., with a 1.5-sec. interstimulus interval. After a practice period with four unrelated words, each of the 15 items of a given list was presented on a Stowe memory drum. Immediately after the list was presented *S* was given either a recall or a recognition test. During any single experimental session, the same test, recall or recognition, was administered for a total of three trials; each trial followed presentation of the same set of 15 items, but in a different random order.

Recall tests. Immediately following list presentation, *S* was asked to say out loud as many of the words on the list as he could remember. He was given 2 min. to do so. The *E* recorded *S*'s responses verbatim. Each subsequent trial was begun at the end of the 2-min. time limit.

Recognition tests. Following list presentation, *S* was given a 90-item printed page which contained all (50 or 54) of the items from the category used to make up the training list, and the remainder consisted of unrelated items (nonset items), such as rabbit, Cadillac, violin, rust, china, oak, etc. The *S* was instructed to check off all of the items which he remembered as having been on the list. Again he was given 2 min. to do so, after which the next trial was begun.

Experimental Design

Half of the Ss in each diagnostic group was given the recognition task first, and 1 wk. later the recall task. The remaining Ss were given the tasks in reverse order. For the first session, half of the Ss in each experimental condition (recall or recognition) was given items from one of the conceptual sets, while the other half was given the alternative conceptual set. One week later, in the second session, when *S* took the other experimental test, he learned the alternate conceptual set. Thus, if *S* was first exposed to a recognition procedure with edible plants as the conceptual set, the following week he would take the recall procedure with states as the conceptual set. Each *S* was given three trials per list, with either

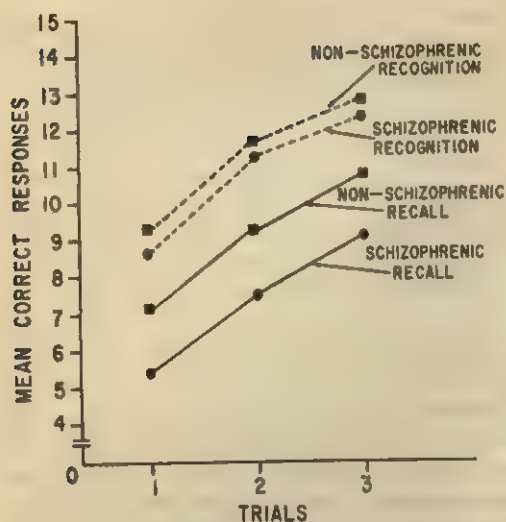


FIG. 1. Mean correct responses over trials.

the recognition or the recall test administered after each trial. The results were analyzed by a within-Ss factorial analysis of variance.

RESULTS

Correct responses. As shown in Figure 1, the results for the recall tests indicated that the schizophrenics produced, on the average, appreciably fewer correct responses than the nonschizophrenics. This difference was significant ($F = 4.05$, $df = 1/56$, $p < .05$). For the recognition tests, however, although the direction of the obtained difference was the same as that found on recall, the difference between the two groups was much smaller and non-significant ($F = .38$, $df = 1/56$).

On each task, the improvement with trials was significant and the same for both groups: (recall, $F = 50.27$, $df = 2/112$, $p < .005$; recognition, $F = 56.32$, $df = 2/112$, $p < .005$). That is, the schizophrenics learned as rapidly as the nonschizophrenics, as measured by both the recall and recognition tests. This can be seen in Figure 1, in the parallel curves for the two groups, with both tasks. Except for the obvious task effect (recognition being superior to recall), no other significant effects occurred, such as those due to different lists (conceptual sets) or order of task presentation.

Thus far the results for correct responses are clearly in line with the predictions generated by the main experimental hypothesis:

The schizophrenics were inferior on the recall task, but approached the nonschizophrenics' accuracy level on the recognition task.^a

Intrusion errors. All overt, incorrect responses were considered intrusion errors and they were examined with respect to two categories: intraset intrusions (items from the same set as those which appeared on the learning list) and extraset intrusions (items unrelated, in terms of conceptual set, to those which occurred on the learning list).

On both tasks there were significant differences in the frequency of intraset intrusions, between the schizophrenics and the nonschizophrenics. That is, the schizophrenics made more intraset intrusions than the nonschizophrenics on recall ($F = 34.61$, $df = 1/56$, $p < .01$), and on recognition ($F = 5.23$, $df = 1/56$, $p < .05$). As indicated in Figure 2, however, the increment in intrusion errors by the schizophrenics was considerably larger on the recall than on the recognition task; the Tasks \times Groups interaction was significant ($F = 8.63$, $df = 1/56$, $p < .01$).

Another feature of these intraset intrusion error results was the finding of a significant Groups \times Trials interaction for the recognition task ($F = 3.33$, $df = 2/112$, $p < .05$). This indicates that with training, the schizophrenics approached the smaller number of recognition intrusions made by the nonschizophrenics. As indicated in Figure 2, by the third trial the schizophrenics' recognition intrusions were almost at the level of the nonschizophrenics.

Extraset intrusions occurred so rarely in either group that detailed statistical comparisons were not feasible. On the recall task

^a The differential power of the recall and recognition tasks to separate the two diagnostic groups might be explained, without recourse to the models outlined earlier in this paper, if the two tasks were not of equivalent difficulty. That is, the difficulty level (average percentage of correct responses) of the recall task may be closer than that of the recognition task to some optimum level for discriminating between the two groups. However, the differences between the two groups, on either task, were found to remain constant from trial to trial even though the scores showed steady intertrial increments over a considerable range (see Figure 1). This suggests that the between-group differences were independent of difficulty level, at least within the limits of the observed ranges.

the mean extraset intrusions for the schizophrenics was .30 on Trial 1, .40 on Trial 2, and .16 on Trial 3; the nonschizophrenics gave no extraset intrusions on any of the recall trials. On the recognition task, the mean extraset intrusions for the schizophrenics was .06 on Trial 1, .10 on Trial 2, and .00 on Trial 3, while the nonschizophrenics produced means of .03 on Trial 1, .00 on Trial 2, and .10 on Trial 3.

One schizophrenic *S* checked off every item on the recognition list. His results were excluded because it was apparent that he was not paying attention to the instructions. In all other instances, *Ss* made so small a proportion of extraset intrusions that it was concluded that they were attending to the task, that is, not responding indiscriminately.

DISCUSSION

The results for correct responses parallel those reported by Cohen and Camhi (1967) for communication accuracy. They found that schizophrenics were significantly inferior to normals in the speaker role, but not in the listener role. Similarly, the present findings for correct responses showed the schizophrenics to be inferior to the nonschizophrenics on the recall task, but not on the recognition task.

Using as a frame of reference the two-stage recall model outlined earlier in this paper, these findings are consistent with the interpretation that schizophrenic *Ss* are defective in the ability to make reliable comparison-stage judgments when the task requires them to integrate comparison with the sampling stage of the two-stage response selection process posited for recall. As predicted, overexclusion (significantly lower correct response scores for the schizophrenic *Ss*) was found on the recall tests, and not on the recognition tests. The results for overinclusion were also in the predicted direction insofar as the increased schizophrenic tendency to make intrusion errors was significantly more prominent on the recall than on the recognition tests.

The fact that some degree of schizophrenic overinclusion was found on the recognition tests, however, does not fully conform to the theory. According to the theory, the schizophrenics were expected to make no more intrusion errors on the recognition tests than

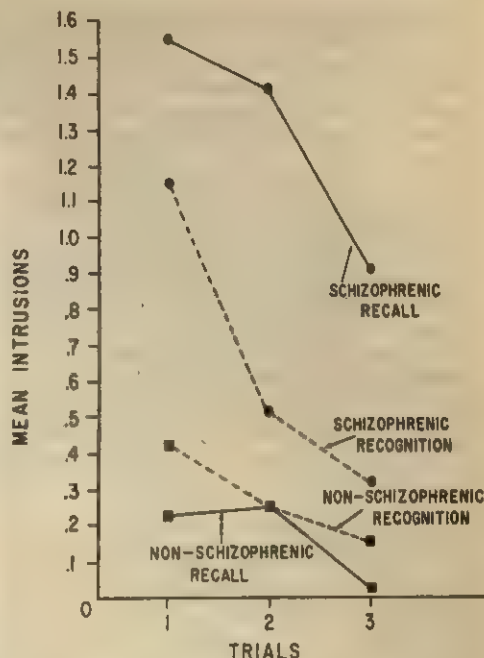


FIG. 2. Mean intrusions over trials.

the nonschizophrenics. A more adequate theory would have to account for the moderate but nonetheless significantly larger number of schizophrenic intrusion errors found on the recognition tests.

A theory recently proposed by Broen and Storms (1966) would seem to account for the prevalence of schizophrenic intrusion errors on both the recognition and recall tasks. Their theory, a variation of one proposed earlier by Mednick (1958), postulates "heightened arousal" in schizophrenics resulting in increased strengths of normally weak, non-dominant responses. The Broen and Storms theory is, however, inconsistent with the decrease in intrusion errors over trials seen in the present experiment on both recall and recognition tasks, and does not account for the differential schizophrenic deficit on the recall versus the recognition tasks.

REFERENCES

- BROEN, W. E., JR., & STORMS, L. H. Lawful disorganization: The process underlying a schizophrenic syndrome. *Psychological Review*, 1966, 73, 265-279.

- CAMERON, N. S. Reasoning, regression and communication in schizophrenics. *Psychological Monographs*, 1938, 50(1, Whole No. 221).
- CHAPMAN, L. J. A reinterpretation of some pathological disturbances in conceptual breadth. *Journal of Abnormal and Social Psychology*, 1961, 62, 514-519.
- COHEN, B. D., & CAMEL, J. Schizophrenic performance in a word communication task. *Journal of Abnormal Psychology*, 1967, 72, 240-246.
- LANG, P. J., & BUSS, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology*, 1965, 70, 77-106.
- MCGHIE, A., & CHAPMAN, J. Disorders of attention and perception in early schizophrenia. *British Journal of Medical Psychology*, 1961, 34, 103-116.
- MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, 55, 316-327.
- PAYNE, R. W., MATUSSEK, P., & GEORGE, E. I. An experimental study of schizophrenic thought disorder. *Journal of Mental Science*, 1959, 195, 627-652.
- ROSENBERG, S., & COHEN, B. D. Referential processes of speakers and listeners. *Psychological Review*, 1966, 73, 208-231.
- SHAKOW, D. Segmental set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry*, 1962, 6, 17-33.
- SPENCE, J. T., GOODSTEIN, L. D., & LAIR, C. V. Rote learning in schizophrenic and normal subjects under positive and negative reinforcement conditions. *Journal of Abnormal Psychology*, 1965, 70, 251-261.
- YATES, A. J. Psychological deficit. *Annual Review of Psychology*, 1966, 17, 111-144.

(Received July 8, 1968)

PREDICTABILITY AND ANXIETY IN SPEECH BY PARENTS OF FEMALE SCHIZOPHRENICS¹

JOSEPH BECKER²

AND

PAUL FINKEL

University of Washington

Kankakee State Hospital, Kankakee, Illinois

Reported anomalies in verbal communication by parents of schizophrenics were investigated. Content structured speech samples from 15 sets of parents of good premorbid (GPM) and poor premorbid (PPM) schizophrenic daughters and nonpsychiatric controls (NPC) were obtained. These samples were analyzed by content area (neutral, aggression, dependency, and sex) for communicability, amount of speech anxiety, and source pathology. Communicativeness was assessed by cloze procedure using male and female college raters. Contrary to prediction, PPM parents tended to have the least amount of speech anxiety and the highest predictability, whereas NPC parents tended to have the highest speech anxiety and lowest predictability. However, raters tended to correctly identify which parents had the most disturbed offspring.

The etiological role of families in psychopathology remains controversial (e.g., Frank, 1965; Rimland, 1966; Zuckerman, 1966). Since Bateson, Jackson, Haley, and Weakland (1956) postulated the "double-bind" hypothesis, distorted parental communication has been increasingly implicated as an important etiological factor in schizophrenia. While the tenability of the double-bind hypothesis has been seriously questioned (Schuham, 1967), subsequent investigations of cognitive and communicative aberrations in parents of schizophrenics have yielded reasonably promising results. Consistently positive indications of deviant thought processes in parents of schizophrenics have been derived from analyses of parental Rorschachs and Thematic Apperception Tests (e.g., Meyer & Karon, 1967; Singer & Wynne, 1963, 1966), and application of these verbal cognitive analyses to object-sorting tasks have likewise yielded encouraging results (Wild, Singer, Rosman, Ricci, & Lidz, 1965).

Investigators of cognitive deviance in parents of schizophrenics contend that vulnerability to schizophrenia is partially mediated by defective symbolic functioning, the core feature of schizophrenic disorders. Parents of schizophrenics are reputed to have disturbed symbolic operations complementary to their disturbed offspring. More specifically, the verbalizations of parents of schizophrenics have been characterized as (a) lacking closure—speech is fragmented, unintelligible, indefinite, contradictory, (b) disrupted—surfeited with irrelevancies, and non sequiturs, and (c) peculiar—idiosyncratic logic, and usage (Singer & Wynne, 1966). These parental anomalies are assumed to impair the development of a capacity for sustained attention, shared meaning, and effective communication within the preschizophrenic child. Singer and Wynne (1966) assert that deviant verbalizations in parents of schizophrenics are not test specific, but are readily identifiable in "almost any verbal material from [schizophrenic] parents [p. 262]," an assumption supported by Morris and Wynne's (1965) postdictive study of family therapy recordings.

The present study investigated the communicativeness of speech from parents of schizophrenics by using cloze procedure (Osgood, 1959, pp. 78-88; Taylor, 1954). This procedure involves substituting blanks for words in verbal messages and requiring receivers to supply the missing words. Cloze procedure is reputed to measure the corre-

¹This investigation was supported in part by a grant from the Psychiatric Training and Research Fund of the Illinois Department of Mental Health, and in part by State of Washington Initiative 171 Funds for Research in Biology and Medicine. Appreciation is expressed to Kashinath D. Patel, Statistical Consultant, University of Washington Primate Center, for guidance in the analyses of these data.

²Requests for reprints should be sent to Joseph Becker, Department of Psychiatry, University of Washington School of Medicine, Seattle, Washington 98105.

spondence between the total language systems of its source and receivers. Osgood (1959) stated that "Cloze works where other methods won't . . . because it is sensitive to just about all of the determinants (semantic, associational, grammatical, and syntactical) which affect encoding by a source and decoding by a receiver [p. 82]." Osgood (1959, p. 88) has speculated that cloze procedure might be a useful tool for assessing thought disturbance. Cloze procedure has discriminated meaningfully in a number of studies related to personality disturbance (e.g., Feldstein & Jaffee, 1963; Osgood & Walker, 1959; Platz & Honigfeld, 1965; Salzinger, Portnoy, & Feldman, 1963).

Neither previous empirical findings with cloze procedure, nor their theoretical rationale, provided a firm basis for predictions. Osgood and Walker (1959), in effect, postulate an inverted curvilinear relation between the drive state of the speech source and the intelligibility of his speech. Moderately high drive maximizes the probability of occurrence of dominant responses within the source's speech hierarchy resulting in greater stereotypy and intelligibility. Extremely high drive pushes multiple responses within the hierarchies to asymptotic strength thus producing blocking, interference, disorganization, and lower cloze scores. Low drive produces flexibility and variability, hence less redundancy, and lower cloze scores. To partially assess the drive state associated with parental speech in the present study, nonlexical speech anxiety levels were determined (Mahl, 1959).

Wynne et al.'s descriptions of the verbalizations of parents of schizophrenics seem analogous to the disorganization Osgood posits for high-drive states. Furthermore, if Wynne et al.'s assumed complementarity between the speech of parents of schizophrenics and their psychotically diagnosed offspring obtains, then the finding of lower cloze scores for schizophrenic sources by Salzinger et al. (1963) would lead to a similar prediction for their parents.

METHOD

Subjects

* Patient criteria required that Ss be cooperative, hospitalized, Caucasian females aged 17-40, non-

retarded or brain damaged, diagnosed as schizophrenic with both biological, native-born parents willing to participate.³ Parents were below age 68, and in reasonably good health. Nonpsychiatric control families (NPC) met the same criteria as patient families, but reported no history of serious personality disturbance. Controls were obtained through local ministers, and received \$10.00. Most patient and control families came from large urban areas (Chicago or Seattle), and were of British, German, or Scandinavian background as judged by surnames.

As previously described (Becker & Siefkes, 1969), parents of patients were split into good and poor premorbid groups (GPM and PPM, respectively) on the basis of their daughters' scoring on the Phillips Premorbid Scale (Farina, Garnezy, Zalusky, & Becker, 1962), and the Social Competence Scale (Phillips & Zigler, 1961).

Median length of patient hospitalization was 2 mo. The sample included 15 pairs of GPM, 15 pairs of PPM, and 15 NPC parents. Parent groups differed significantly ($p < .05$) in age, education, and estimated IQ (Wechsler Adult Intelligence Scale, WAIS, Vocabulary subtest), but not in socioeconomic class (Hollingshead, 1957). The PPM parents had less education and lower vocabulary scores than GPM and NPC parents.

Procedure

Separately and in private, each parent was asked to tape-record 4 min. of speech in each of four content areas. The first area was a relatively neutral unstructured one; parents spoke on any aspect of family life they cared to. Instructions were:

I'm going to ask you to do something that may be a little difficult, but I'd like you to do it as well as you can. We're interested in what sort of ideas occur to people about their family life without thinking it over. For four minutes, I'd like you to speak whatever thoughts come to your mind about your family life. I will listen and jot down some notes as you speak. I would like for you to do all the talking during this four minutes. In order not to influence your stream of thoughts, I will try to have as little facial expression as possible.

The other three content areas (dependency, aggression, and sex) related to childhood socialization practices with their schizophrenic (or control) child. Order of presentation of the socialization content areas was counterbalanced. As a preface to the socialization content areas the following introduction was given:

Okay, this time I'd like to hear you talk for about four minutes about how you handled some more

³ Some parents of the schizophrenic patients had been previously hospitalized with schizophrenic diagnoses. None of the parents was overtly psychotic when this study occurred, and none was excluded on this basis.

general problems that all parents have in raising their children. Let's take the problem of how you dealt with [first name of patient or control] during her first few years of life when you were [dependency] trying to get her to show her wishes for help, attention, and affection in an acceptable way, [sex] trying to teach her about sexual matters such as sex play with either herself or other children, and about going around with clothes on, [aggression] trying to get her to control her anger, destructiveness, and fighting.

After each 4-min. section, *E* would say, "Alright, fine. Now how did you deal with [daughter's name] when you were . . ."

Preparation of Cloze Protocols

The first 150 words of each speech sample were transcribed with every fifth word deleted and replaced by a line of standard length. All nonlexical verbal expressions were included in the protocol but not counted as words. Since previous work (Becker, Tatsuoka, & Carlson, 1965) had suggested that sex of rater might interact with sex of speech source, two sets of 30 protocol booklets, each containing 12 protocols, were compiled. One set of 30 booklets was administered to male raters, the other set to female raters. Speech protocols were randomly assigned to booklets within sets except that half of the protocols in each booklet came from each sexed parent.

Raters and Instructions

Basic raters were 30 male and 30 female University of Washington enrollees in introductory psychology who were drawn randomly from the *S* pool. A female graduate research assistant met with student raters in groups of 8-12. Rater groups were instructed as follows:

This study is part of a larger one concerned with the language of parents who have or do not have emotionally disturbed offspring. Your booklet contains twelve pages. On each page there is a sample of speech from a different parent. Approximately every fifth word in each sample has been deleted and replaced by a blank of uniform length. Your task is to guess what the missing word is and to write your guess in the blank space. No blanks were substituted for proper nouns, numbers, or exclamations. All proper nouns have been changed to preserve confidentiality. Contractions like "don't" have been counted as one word in substituting blanks. A series of dashes in parentheses (---) indicate that the tape recording of the parent's speech was unintelligible at that point. Each dash indicates about one syllable of missing speech.

Before attempting to fill in the blanks on a page, read the entire page first. This will provide a feel for the person's language habits. You will have five minutes for each page which is quite ample. It is very important that *every* blank be filled in.

One final request—at the bottom of each sheet you will notice a rating scale like so:

Please check the one category that you think is most likely correct:

This parent has a severely disturbed child._____

This parent has a moderately disturbed child._____

This parent probably does not have a disturbed child._____

Please fill in this scale for *each* sheet.

When you have completed each sheet, please do *not* go on to the next one until asked to do so by the investigator.

For a protocol to be accepted as suitably rated, at least 12 of the 15 blanks had to be filled in; and for a booklet to be accepted, 10 of the 12 protocols had to meet the criterion. These criteria were adopted to help ensure adequate rater motivation. Three booklets and 7 additional protocols failed to meet the criteria. These "failures" seemed randomly distributed over the protocols. Additional raters were obtained to fulfill the requisite number of rated protocols by each sex.

RESULTS

Cloze Analyses

Arcsin transformations were made on the raw score cloze proportions. To determine whether covariance adjustments were indicated, correlations were determined for each sexed parent between each of the supplementary variables on which the parents differed (age, education, and WAIS Vocabulary), and the cloze scores for each of the speech content areas. Since only 2 correlations of 48 were significant ($p < .05$), covariance adjustments were not used. Likewise, Mahl speech anxiety scores were unrelated to cloze scores. Data for the cloze scores of each content area were analyzed separately by partially nested repeated-measures analyses of variance. This design permitted testing of group effects, and Sex of Parent \times Group interactions. Comparisons of ratings by male versus female raters and of mothers versus fathers were made by *t* tests. Results of the analyses of variance are presented in Table 1. Mean cloze scores are depicted in Figure 1.

Overall findings provided no support for the predicted low cloze scores of PPM parents. On the contrary, PPM mothers consistently obtained the *highest* cloze scores and NPC mothers the lowest. Significant interactions occurred for all socialization content areas. These interactions were chiefly due to the

TABLE 1
ANALYSES OF VARIANCE OF CLOZE SCORES

Source	df	Content area							
		Neutral		Aggression		Dependency		Sex	
		MS	F	MS	F	MS	F	MS	F
Groups (A)	2	2.43	— ^a	320.01	4.97 [*]	130.77	1.86 ^a	21.78	— ^a
Families within groups (B in A)	42	63.43		64.36		70.28		94.65	
No. parents within family (C in B)	45	58.63	1.33 ^b	56.06	3.23 ^{b***}	77.83	2.85 ^{b***}	111.76	2.08 ^{b**}
No. raters within parent (D in C)	90	44.17		17.35		27.32		53.80	

^a MS for A divided by B in A.

^b MS for C in B divided by D in C.

* $p < .025$.

** $p < .01$.

*** $p < .001$.

relatively high cloze scores of PPM mothers and the relatively low scores of their husbands. Male and female raters' scores did not differ for either parent in any content area.

Neutral speech. There were no significant group differences.

Aggression speech. Duncan multiple-range tests yielded lower cloze scores for combined NPC parents than for GPM ($p < .05$) and PPM parents ($p < .01$). Among the mothers' groups, NPC groups were lower than PPM groups ($p < .05$); and among the fathers' groups, GPMs were higher than NPC

($p < .05$) and PPM groups ($p < .05$). Although the interaction ($p < .001$) is largely due to differences between the PPM mother and father groups, these two groups did not differ significantly between themselves ($t = 1.11$, $df = 14$).

Dependency speech. The trends of results were similar to those for aggression with the interaction ($p < .01$) due mainly to the divergence between PPM mothers and fathers ($t = 2.15$, $df = 14$, $p < .05$). There were no differences among the mother groups or among the father groups by Duncan multiple-range test.

Sex speech. While there was no diagnostic group difference in cloze scores for sex, mothers overall had higher scores ($t = 2.41$, $df = 88$, $p < .05$) than fathers. The interaction ($p < .01$) was due to the tendency of NPC and PPM mothers to have higher scores than NPC and PPM fathers, whereas GPM parents were similar.

Mahl Analyses

To determine interrater reliability, 12 of the speech protocols were randomly selected and rescored blindly by an independent rater. A Spearman rank correlation rho of .95 was obtained. Intercorrelations between the Mahl and concomitant variables yielded no significant correlations ($p < .05$), hence no covariance adjustments were used with the two-factor mixed design with repeated measures (Groups \times Parents). Means and standard deviations of the groups for each content area are presented in Table 2. No significant

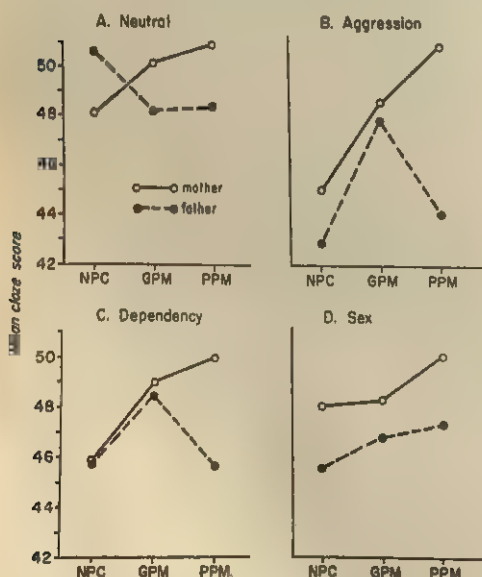


FIG. 1. Mean cloze scores for mothers and fathers of NPC, GPM, and PPM daughters.

within-Ss differences were obtained either between mothers and fathers or as an interaction between parents and groups. The PPM parents, especially the fathers, consistently tended to show *least* speech anxiety and NPC parents most. Significant between-group differences were found for neutral speech ($F = 6.03$, $df = 2/42$, $p < .01$) and dependency speech ($F = 3.38$, $df = 2/42$, $p < .05$).

Neutral speech. Both NPC and GPM parents evidenced more speech anxiety than the PPM group ($p < .01$) by Newman-Keuls test. Whereas the mother groups did not differ among themselves, the father groups did, with both NPC and GPM fathers showing more speech anxiety than PPM fathers ($p < .005$ and $< .05$, respectively) by Duncan multiple-range test.

Aggression speech. A Newman-Keuls comparison of parent groups yielded lower speech anxiety for PPM than NPC groups ($p < .05$). Again mothers did not differ, but NPC and GPM fathers had more speech anxiety than PPM fathers ($p < .005$ and $< .01$, respectively) by Duncan multiple-range test.

Dependency speech. Nonpsychiatric control parents were more anxious than PPM parents ($p < .05$), but separate comparisons among mother groups and among father groups yielded no differences.

Sex speech. There were no differences between groups.

Severity of Illness Ratings

Despite the PPM parents' trend toward higher cloze scores and lower speech anxiety, raters tended to infer correctly that PPM parents had reared the most disturbed children. Neither male versus female raters' nor mothers' versus fathers' scores differed. Partially nested repeated measures analyses of variance were used. For neutral speech, groups differed ($F = 4.16$, $df = 2/42$, $p < .025$) with combined NPC groups rated as having less disturbed children than either GPM or PPM parents ($p < .05$); the latter two groups did not differ. The same trend appeared for aggressive and dependency speech, but were not statistically significant. Statistically significant but small and inconsistently patterned interactions occurred for Sex of Parent \times Diag-

TABLE 2
MAHL ANXIETY MEANS AND STANDARD DEVIATIONS

Content area	Mothers		Fathers	
	\bar{X}	SD	\bar{X}	SD
Neutral				
NPC	10.05	3.89	11.90	3.23
GPM	10.92	2.77	10.57	3.32
PPM	9.57	2.89	8.07	2.82
Aggression				
NPC	11.10	4.88	13.75	3.73
GPM	11.64	4.13	10.75	2.10
PPM	10.67	2.92	9.79	2.49
Dependency				
NPC	12.75	3.14	12.87	4.99
GPM	12.63	2.81	11.47	3.71
PPM	10.63	3.80	10.57	2.99
Sex				
NPC	12.50	3.81	13.50	2.88
GPM	12.47	2.68	12.73	3.63
PPM	11.53	4.49	11.83	3.17

Note.—NPC = nonpsychiatric controls; GPM = good premorbid; PPM = poor premorbid.

nistic Group for both aggressive ($F = 1.64$, $df = 45/90$, $p < .05$) and dependency speech ($F = 1.41$, $df = 45/90$, $p < .05$).

DISCUSSION

The findings of relatively high speech predictability in PPM mothers and low anxiety in PPM mothers and fathers were contrary to expectation. Probably the high cloze scores of PPM mothers were due to considerable stereotypy which Osgood associates with high drive state (D) as construed within the Spence-Hull framework. The findings of the present study indicate that either Mahl's speech anxiety measure is not coordinate with D or that Osgood's theory requires modification.

Post hoc, earlier work of Wynne's (Wynne, Ryckoff, Day, & Hirsch, 1958) group on the concept of pseudomutuality could be reconciled with an expectancy of higher cloze scores in PPM parents. They emphasize their lack of spontaneity and novelty, their excessive use of bland indiscriminate approval, their massive use of denial, and constricted purview. Further support for greater constriction and stereotypy in PPM mothers than fathers derives from Becker and Siefkes' (1969) finding that the PPM mothers reported a narrower range of disciplinary measures than any of the other parental groups.

Likewise, the low anxiety of PPM parents can be rationalized. Lidz, Fleck, and Cornelison (1965) in summarizing observations on parents of female schizophrenics emphasize the "aloof" qualities of their fathers (pp. 104-105) and mothers (p. 264). Baxter, Becker, and Hooks (1963) reported greater use of denial than obsessional mechanisms in PPM parents. Recent work with Byrne's (1964) Repression-Sensitization (R-S) scale suggests that less overt anxiety is associated with denial and repression than with obsessiveness.

Also, it is possible that PPM parents, perhaps especially the mothers, have been queried more about their socialization practices, and have thought more about them. Hence they may have been more rehearsed and less anxious.

Finally, the data in the present study were obtained under somewhat more structured conditions than Wynne et al.'s conjoint family therapy sessions or projective testing. Degree of structuredness may be a critical determinant of the coherence and cohesiveness of PPM parents' speech and thought.

It appears that the generality of Wynne et al.'s findings must be viewed with caution. The tendency of raters in the present study to correctly identify PPMs as parents of more severely disturbed children is, of course, congruent with Wynne et al.'s findings. Increased concern with varied measures of assessment and further delineation of dimensions of thought disorders are clearly indicated.

The most valid generalization to be drawn from the present findings is that interesting and unexpected results have been obtained which require further explication.

REFERENCES

- BATESON, G., JACKSON, D. D., HALEY, J., & WEAKLAND, J. Toward a theory of schizophrenia. *Behavioral Science*, 1956, 1, 251-264.
- BAXTER, J. C., BECKER, J., & HOOKS, W. Defensive style in the families of schizophrenics and controls. *Journal of Abnormal and Social Psychology*, 1963, 66, 512-518.
- BECKER, J., & SIEFKES, H. Parental dominance, conflict, and disciplinary coerciveness in families of female schizophrenics. *Journal of Abnormal Psychology*, 1969, 74, 193-198.
- BECKER, J., TATSUOKA, M. M., & CARLSON, A. R. The communicative value of parental speech in families with disturbed children. *Journal of Nervous and Mental Disease*, 1965, 141, 359-364.
- BYRNE, D. Repression-sensitization as a dimension of personality. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1964.
- FARINA, A., GARMETZ, N., ZALUSKY, M., & BECKER, J. Premorbid behavior and prognosis in female schizophrenic patients. *Journal of Consulting Psychology*, 1962, 26, 56-60.
- FELDSTEIN, S., & JAFFE, J. Language predictability as a function of psychotherapeutic interaction. *Journal of Consulting Psychology*, 1963, 27, 123-126.
- FRANK, G. H. The role of the family in the development of psychopathology. *Psychological Bulletin*, 1965, 64, 192-205.
- HOLLINGSHEAD, A. B. *Two-factor index of social position*. Unpublished manuscript. Available from the author, Yale University, 1957.
- LIDZ, T., FLECK, S., & CORNELISON, A. R. *Schizophrenia and the family*. New York: International Universities Press, 1965.
- MAHL, G. F. Exploring emotional states by content analysis. In I. De Sola Pool (Ed.), *Trends in content analysis*. Urbana: University of Illinois Press, 1959.
- MEYER, R. G., & KARON, B. P. The schizophrenogenic mother concept and the TAT. *Psychiatry*, 1967, 30, 173-180.
- MORRIS, G. O., & WYNNE, L. C. Schizophrenic offspring and parental styles of communication. *Psychiatry*, 1965, 28, 19-44.
- OSGOOD, C. E. The presentational model and relevant research methods. In I. De Sola Pool (Ed.), *Trends in content analysis*. Urbana: University of Illinois Press, 1959.
- OSGOOD, C. E., & WALKER, E. G. Motivation and language behavior: A content analysis of suicide notes. *Journal of Abnormal and Social Psychology*, 1959, 59, 58-67.
- PHILLIPS, L., & ZIGLER, E. Social competence: The action-thought parameter and vicariousness in normal and pathological behaviors. *Journal of Abnormal and Social Psychology*, 1961, 63, 137-146.
- PLATZ, A., & HONIGFELD, G. Some effects of anxiety on the intelligibility of verbal communication in psychotherapy. *Journal of Personality and Social Psychology*, 1965, 2, 122-125.
- RIMLAND, B. Psychiatry overextended. *Science*, 1966, 154, 1935.
- SALZINGER, K., PORTNOY, S., & FELDMAN, R. S. Verbal behavior of schizophrenic and normal subjects. *Annals of the New York Academy of Sciences*, 1964.
- SCHUHAM, A. I. The double-bind hypothesis a decade later. *Psychological Bulletin*, 1967, 68, 409-417.
- SINGER, M. T., & WYNNE, L. C. Differentiating characteristics of parents of childhood schizo-

- phrenics, childhood neurotics, and young adult schizophrenics. *American Journal of Psychiatry*, 1963, 120, 234-243.
- SINGER, M. T., & WYNNE, L. C. Principles for scoring communication defects and deviances in parents of schizophrenics: Rorschach and TAT scoring manuals. *Psychiatry*, 1966, 29, 260-268.
- TAYLOR, W. L. Application of cloze and entropy measures to the study of contextual constraints in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.
- WILD, C., SINGER, M., ROSMAN, B., RICCI, J., & LIDZ, T. Measuring disordered styles of thinking. *Archives of General Psychiatry*, 1965, 13, 471-477.
- WYNNE, L. C., RYCKOFF, I. M., DAY, J., & HIRSCH, S. I. Pseudo-mutuality in the family relations of schizophrenics. *Psychiatry*, 1958, 21, 205-220.
- ZUCKERMAN, M. Save the pieces! A note on "The role of the family in the development of psychopathology." *Psychological Bulletin*, 1966, 66, 78-81.

(Received July 10, 1968)

Manuscripts Accepted for Publication in the
Journal of Abnormal Psychology

- Habituation to a Loud Sound as a Function of Manifest Anxiety: Seymour Epstein* and Walter D. Fenz: Hohfluh, Hasilberg, 6082 BO, Switzerland.
- Dr. Jekyll and Mr. High: Multidimensional Scaling of Alcoholics' Self-Evaluations: John T. Partington*: Department of Psychology, University of Western Ontario, London, Canada.
- Dimensions of Conceptual Behavior: A. B. Silverstein*, Richard E. McLain, and Philip J. Mohan: Acting Chief of Research, Pacific State Hospital, Box 100, Pomona, California 91766.
- Evaluative Feedback and Aspiration Statements of Schizophrenics and Normals: C. Clifford Atkisson and James M. Anker*: Department of Psychology, University of Tennessee, 1303 West Cumberland Avenue, Knoxville, Tennessee 37916.
- Deterioration Effect: A Reply to Braucht: Allen E. Bergin*: Department of Psychology, Teachers College, Columbia University, New York, New York 10027.
- Deterioration Effect: A Reply to Bergin: G. Nicholas Braucht*: Manpower Studies, Colorado State University, 50 West 5th Avenue, Denver, Colorado 80204.
- Comment on "High School Yearbooks: A Nonreactive Measure of Social Isolation in Graduates Who Later Became Schizophrenic": J. Conrad Schwarz*: Department of Psychology, Syracuse University, 331 Huntington Hall, Syracuse, New York 13210.
- Reliability of Chronicity as Rated from Case History Summaries: George A. Clum* and George R. Bowen: Department of the Navy, Navy Medical Neuropsychiatric Research Unit, San Diego, California 92152.
- Early Infantile Autism: A Methodological Critique: John C. Masters* and Dick E. Miller: Institute of Child Development, University of Minnesota, Minneapolis, Minnesota 14.
- Vocal Isolation: Effects of Occluding Auditory Feedback from One's Own Voice: George S. Klein and David L. Wolitzky*: Research Center for Mental Health, New York University, 4 Washington Place, Washington Square, New York 10003.
- Relationship between Initial Competence and Ability to Profit from Cues in Brain-Damaged Individuals: Yehuda Ben-Yishay*, Leonard Diller, Louis Gerstman, and Wayne Gordon: Institute of Rehabilitation Medicine, New York University Medical Center, 400 East 34th Street, New York, New York 10016.
- Fate of an Uncompleted Posthypnotic Suggestion: Edgar P. Nace and Martin T. Orne*: Institute of the Pennsylvania Hospital, Unit for Experimental Psychiatry, 111 49th Street, Philadelphia, Pennsylvania 19139.
- Illicit LSD Users: Their Personality Characteristics and Psychopathology: Reginald G. Smart* and Dianne Jones: Addiction Research Foundation, 344 Bloor Street West, Toronto 4, Ontario.

*Asterisk indicates author for whom address is supplied.

CONSTRUCT VALIDITY FOR THE EVALUATION OF THERAPY OUTCOMES¹

MERTON S. KRAUSE²

Institute for Juvenile Research, Chicago, Illinois

Evaluation is a matter of judging conformity to certain values and the judgment of how healthy or diseased a person is is an evaluation. The first step in doing outcome evaluation studies of psychotherapy is, therefore, to explicitly choose a reference group or public and find out what its relevant values are. Patients, clients, sponsors, and therapists constitute four such publics with somewhat different values. Probable value criteria for these publics were proposed and the problems of defining and collecting valid data for these criteria were discussed.

The impasse between clinicians and researchers over the value of psychotherapy exists primarily because of the widespread misconception that valid evaluative measures of outcome can be discovered in empirical studies of or experience with therapy or can be derived from theories of personality. Instead, they must be derived from or proved by the values held by certain groups of persons. Why this is so, what these values may be, and how outcomes must therefore be evaluated are the topics considered in this paper.

VALUES AND PUBLICS

There can be no evaluation without prior standards for what is good or what is better. Such standards cannot be logically deduced (or invented) without a priori value propositions to deduce them from. Likewise, they cannot be empirically discovered without a priori criteria, that is, value propositions, for recognizing which differences in outcome between treatments are relevant to evaluation and the degree to which each criterion is important (see, e.g., Margolis, 1966, pp. 56-66; Wootton, 1959, pp. 203-226). Since an evaluation is always done for some audience or public, it can only be convincing if it is done

in terms of that public's relevant values. If such a public has no explicit relevant values or no adequate consensus on what these values are, no evaluation done for it can probably be decisive or even acceptable. The achievement of an adequate value consensus is a social process wherein certain values come to be widely held through persuasion, negotiation, legislation, or the development of dogmatic traditions. New factual discoveries, however, may indicate to some members of a public that a certain value commitment is inadequate, in that it is too narrow or internally contradictory, for example, but they cannot imply what it should be.

The evaluation of therapies intended to enhance health or reduce disease in humans must be done in terms of the values which define these human conditions.³ The public whose values must be satisfied in any particular case includes those persons in whose interest the therapy is performed or provided.

¹ This work was supported in part under Public Health Service Mental Health Project Grant MH 00764 from the National Institute of Mental Health. It was facilitated by the advice and criticism of James Breedlove, Jay Hirsch, Kenneth Howard, Mary Richards, William Seeman, and Kathleen Zechmeister.

² Requests for reprints should be sent to the author, Institute for Juvenile Research, 232 East Ohio Street, Chicago, Illinois 60611.

³ The concept of disease is used quite broadly to include all sorts of unhealthiness. Thus, injury and malaise are included just as are infectious diseases and neurosis. Even character disorders, disabilities, and retardations are covered. The values that generically define disease are ultimately cultural or social rather than medical or other professional ones (e.g., Dubos, 1961, p. 214; Jaspers, 1962, pp. 779-788; Mechanic, 1962). The matter could be relegated to some profession, but, if it is, such a delegation of authority should first be well and openly considered. To be judged ill is often tantamount to some loss of freedom and other rights, especially if mental illness is imputed. The author does not agree with Szasz (1961, pp. 39-45) that the answer lies in limiting the definition of illness to physiochemical disorders. Instead, a full scale analysis of the concept of illness is in order.

This means that the values of (a) the persons receiving the therapy (the patients), (b) the persons whose complaints will be remedied by the therapy (the clients), and (c) the persons whose planning and material support is responsible for the provision of the therapy (the sponsors) are pertinent to the therapy's evaluation. In some cases these are all the same person, such as the affluent self-referral for psychotherapy. In other cases they may all be different people, as in a school referral of a fire-setter to a public child guidance agency. But all the variations are possible, for example, that of a housewife who brings in her employed alcoholic husband to a private family counseling agency.

The only other obviously interested parties to therapy transactions are (d) the professionals who "perform" them, but these professionals are almost always selling their services as means to someone else's ends as well as applying their own professional norms concerning means and ends. Therefore, as responsible vendors, they should subscribe to the values of their sponsors; as honest contractors, to the values of their clients; and, as humanitarians, to the values of their patients. Professional norms do further complicate the therapists' values but can be considered separately where different from sponsor, client, or patient values (as in the judgment the operation was a success but the patient died). It may be that the outcome of a therapy case could be evaluated differently in terms of the different values of these four publics. And it is not at all clear that the values of any one public, whatever these may be, are superior to or more relevant than the values of any other. Thus, issues such as whether or not "mental illness" is disease are cultural issues, not merely scientific ones, and so are settled ultimately by whatever means produce a relevant value consensus.

Health and disease are both open concepts, that is, only partially defined. It is obvious that they are extensionally open (that all their instances are not known) because new cases of disease are regularly discovered, and even altogether new diseases and degrees of health are occasionally identified.⁴ It is not

quite so obvious that they are also intensionally open concepts, that is, that the criteria for distinguishing a diseased from a healthy person cannot be completely stated to everyone's satisfaction. There are, however, some definitional constraints that are, we believe, generally respected. Since this is a methodological analysis, rather than an empirical survey of definitions in use, this purportedly consensual theme is developed.

HEALTH AND DISEASE

Whatever other characteristics it may have, every disease involves or portends bad consequences for at least the person who has the disease. A *sine qua non* of disease is the presence of or disposition to disvalued (i.e., bad) experiences of, acts by, or physical events in the person which tend to incapacitate him for the role performances expected of him by the relevant publics (see Parsons, 1964, pp. 258-262). Pain is the paradigmatic experiential event; destruction, the paradigmatic behavioral event; and loss of, or inadequate, form or function, the paradigmatic physical or characterological events, respectively. These disvalued and incapacitating events are the primary characteristics of all diseases, mental and physical, from their relation to which the rest of any disease's characteristics acquire their relevance to that disease. These primary characteristics are weakly disjunctively necessary conditions for the presence of manifest disease. Although at least one of them is necessary, no specific combination of these three characteristics is generally sufficient to define a state of disease. This is so, in part, because tolerances for pain, destruction, and losses of form and function vary, and because tolerances for different degrees of incapacitation in different roles vary as well. However, it is also due in part to the priority of certain alternative explanations for some instances of these primary characteristics, for example, overexertion, immaturity, or aging.

One reason for these primary disvalued characteristics being necessary to the definition of any disease is the evaluative nature of

⁴ Thus, there are those who believe that a client's failure to realize his potential, to actualize, develop,

and express himself is an index of mental illness (e.g., Maslow, 1962, pp. 141-174).

the disease concept itself.⁵ To diagnose a person as having a disease is tantamount to a decision to take therapeutic action (conditional upon the availability of the resources for doing so, upon competing claims on these resources, and upon the probable net advantage to patients, clients, or sponsors involved). Thus, the identification of an instance or case of disease by a medical or other therapeutic professional in the line of duty (i.e., in response to a bona fide client) is not a mere perceptual or cognitive achievement. It is also, properly, often a cue to some more or less remedial response, *because* a state of disease is always a disvalued state of any person.

Within the context of disvalued and incapacitating personal events, as the class of (necessary) disease-defining events, the several possible publics for evaluative research can differ in their values. They can also differ in their skill at recognizing dispositions to, if not the presence of, these primary characteristics. It is here that therapy professionals may properly claim some technical advantage, though their degree of advantage is an empirical matter. Evaluative research, however, need not endow these prognostications with any privileged status so long as extended or recurrent contact with (former) patients and clients is feasible.

MENTAL HEALTH STATUS

Some disease, broadly conceived, seems to be caused at least in part by mental or personality characteristics of the patient. This kind of "disease" has been taken by the "mental health" professions as their responsibility. In no way, however, should the logical

problem of evaluating the outcome of the psychosocial therapies used by these professionals differ from that of evaluating the outcome of any treatment of any disease (see Margolis, 1966, pp. 71-82, on this point). The previous discussion shall merely be extended to more nearly depict operational assessment procedures for mental health (or disease) status.

Besides physical pain there are equally disvalued moods or affects, such as anxiety, despair, and guilt, which can be quite incapacitating. The destructive acts to self or others which can indicate disease are more indigenous to mental than to physical illness, though here it is their irrational quality which is often stressed. The inadequate attainment or the impairment of form and function are, in this realm of disease, usually in social or mental affairs, rather than in physical structures. The incapacitation for role performance, however, can be as severe as in any other realm of disease. How unhealthy, ill, sick, or diseased a person is, and, consequently, how much better or worse he is, can, of course, only be determined from some value position (although his prognosis, in terms of whatever normative framework is adopted, is in part a matter of scientific knowledge).

Each *patient* has his own tolerances for various moods and affects, his special personal and environmental resistances to ostensibly destructive behavior, and his particular personal and environmental tolerances for inadequate or reduced mental and social form and function. Furthermore, the roles which he must, or hopes to, undertake and the demands they make upon him are also specific to his life. Therefore, any evaluation of his therapy from *his* standpoint must concern changes in experiences, acts, events, and capacities as judged by him to be better or worse on balance.

A *client* who is not a patient (i.e., someone who refers a patient at least in part because of his bad effect on their lives or jobs) can make similar kinds of appraisals but only from his own standpoint. These appraisals by collateral persons can diverge considerably from those of the patient himself. Their recognition and valuation of the patient's moods, destructive acts, functioning, and capacities

⁵ Therefore, statistical abnormality per se is insufficient as a defining property of disease, because valued states of the person may also be rare. It becomes absurd as a defining property of seriousness of disease or degree of unhealthiness, because it makes more frequent diseases ipso facto less serious. Thus, without arbitrarily specifying some reference frequency or population, burgeoning epidemics make the individual patients improve, and effective preventive and curative procedures make them deteriorate. Only by affecting values can abnormality influence the definition of disease, but it has no privileged status among the many factors that can influence values.

are quite independent of his. Perhaps his reports of his own affective experiences are better than theirs, but this does not hold for impressions of the effectiveness, finesse, efficiency, or public consequences of his role performances. They may see problems for themselves, created by the patient, which the patient either does not see or interprets differently.

A *sponsor*, who is neither a patient nor client, may, as an enlightened humanitarian, share a concern with patients for their experiences, self-destructiveness, and the functioning and capacities they value specifically. He may also be concerned about the problems that patients create for various collateral persons. However, he may, further, be concerned with law and order, public decency, responsible citizenship, escapism, economic self-reliance, etc., in ways which patients and clients do not see as relevant to therapy. (What, for example, does government foresee as the desirable consequences of public mental health centers?)

The mental health *professionals* are undoubtedly concerned to optimize the patient's outcome from his standpoint (even, sometimes, at the expense of altering that standpoint a bit) and from the standpoint of certain other clients as well. The parents in child guidance cases, the children of a senile parent, the child's analyst who has referred the parents for counseling (but not so often the wife of an alcoholic or the dependent mother of a bachelor), may find the therapist sympathetic to their value position. Therapists probably are also relatively aloof to purely sponsor preferences. Most of all, therapists rely on their own special standpoint for judging mental health status. Unfortunately, what their value position is varies considerably both within and between the therapy professions (e.g., see Offer & Sabshin, 1966), so it is discussed here in quite broad terms.

The major novelty in the therapist's evaluative standpoint is its inclusion of prognostic indicators. Thus, judgments about ego strength, object relations, superego severity, opportunities for impulse gratification, skill repertoire, family dynamics, environmental stresses, etc., influence their evaluations of

mental health status. The only justification for such a use of prognostic indicators is in the degree to which they actually do tend to predict the state of the several primary characteristics of disease already discussed. Another novelty of the therapists' evaluative standpoint, before we pursue the first more deeply, is its usual restriction to data based on observations made in the office, clinic, or hospital, environments quite foreign to the life for which the patient is being treated.

SPECIFIC DISEASES

Among the persons who are considered diseased (sick, ill, or injured) by a medical or paramedical professional, discriminations are made regarding what disease or diseases they "have." In our society the recognition and nonrecognition of disease by medical professionals has acquired a special legitimacy, although their special experience and skills justify such legitimacy only for their discriminations of *specific* diseases. These latter are defined most securely when they are defined genotypically, that is, when their course, etiology, prevention, and cure are understood (see Temkin, 1963). Extensive scientific research is necessary to achieve such an understanding. Short of such a genotypic definition, and prerequisite to it, is phenotypic definition, which depicts the configuration and covariation of the general and primary disease characteristics with any others. Such reliable clusters of characteristics are the bases upon which discriminable kinds of diseases are recognized and their genotypic definitions sought. The crucial point, however, is that the criterion indicators of disease in general are the primary characteristics previously discussed, whereas the criterion differential indicators of specific diseases may be anything. What these latter properly are is clearly an empirical rather than a value or stipulative matter, but it is quite irrelevant to evaluating health status unless these differential diagnostic indicators are well-evidenced predictors of future stability or change in the primary characteristics of disease. Such demonstrated prognostic utility is practically unknown in the archives of mental health research. Therefore, evaluative studies of psychosocial ther-

apy should not rely upon the inadequately predictively validated prognostic indicators.

The exception to this rule quite properly occurs when the public for which an evaluative study is performed is that of therapists. (The happenstance that outcome studies have been predominantly, if not exclusively, done by and for therapists and allied professionals may make this logical exception seem more the rule.) If the goals of therapy are cast in terms of such prognostic indicators, then it is of interest to study when and how these goals are attained. Such technical or professional studies, or those devoted to theoretical or predictive validity, however, ought not be mistaken for true evaluative studies of mental health status outcomes. Likewise, studies concerned with the stability or change of a specific disease in patients must not be equated with studies of general health status, because one disease may improve and yet the patient may be more unhealthy. Since the seriousness of any two specific diseases can logically be negatively correlated within patients and since all disease has not been classified into a closed set of specific diseases, the course of any one specific disease is no sure determinant of health status.

CONSTRUCT VALIDITY*

The first problem of evaluative data is their semantic construct validity, that is, to what extent they are relevant to or can reflect health status for a given public. The line of argument set forth so far was intended to explicate the concepts of health and illness. Although the concept of *mental* health has been used, it is certainly not clear how to distinguish mental health from physical health unless, for example, any genotypically defined physical diseases which are demonstrably unresponsive to psychosocial therapies be excluded. But what could be the purpose of such a distinction for evaluative research, except to create two subprofessions out of one. For the *primary* criteria of outcome evaluation no different background or training is necessary to do research on treatments for tuberculosis or fractures as to do it on schizo-

phrenia, homosexuality, or low back pains. And these primary criteria are, in addition to cost, what matter in treatment evaluation. Since the explication includes considerations popular among therapy professionals, one expects that it should be more or less acceptable to them, at least as far as it goes. Nevertheless, there can be no rational evaluation without specifying the public for which it is expressly designed, and so empirical research on that public's value position is prerequisite.

The second problem of evaluative data is that their public may be committed to certain theoretical propositions so firmly that outcome findings contrary to these would be considered invalid. Thus, psychoanalysts might find it implausible that good analysands would not be healthier on the average after 2 years of apparently successful analysis, and client-centered therapists might feel the same about a similar exposure to perceived unconditional positive regard, empathic understanding, and congruence in the relationship, etc. The only insurance an evaluative researcher can have against such inconclusiveness is to achieve and maintain his chosen public's faith in the validity of his data. It must be his goal to collect sufficiently convincing data to jar the prior convictions of even the most theoretically committed of his public, should disconfirmative results turn up. That is exactly the upper bound of how much validity he needs.

A usual line of attack on the validity of data, and so the third problem for an evaluator, concerns technological construct validity. There are always possible biasing factors, not colinear with actual mental health status, which may have influenced the data. Respondent, test taker, interviewer, test administrator, and *E* biases have been demonstrated, but may be stochastically equated between experimental and control groups. What cannot be avoided through better assessment instruments must be rendered irrelevant by better experiment design. For this reason if no other, control groups are generally necessary in evaluative research.

Although most "evaluative" work is done for therapists, and some ostensibly for (especially Federal) sponsors, it is conceivable that consumers' organizations will someday be strong enough in this country to purchase

* The methodological grounds for this section are discussed in Krause (1967).

their own research for patients' and clients' standpoints. In the meantime we must be content with exploring logical possibilities and encouraging the development of countervailing forces in the market of health services.

PRACTICAL EVALUATION⁷

Once a random or adequately matched assignment of patients between a control "treatment" and the therapy being evaluated has been achieved, the conditions for some evaluation are present. In whatever terms are acceptable to the evaluator's public, the roles and standards of performance of the patients must be assessed. Instrument development must be based upon empirical research primarily *on the relevant public*. There are no scientific standards for the content of evaluative assessments, unless "well-evidenced" predictors of future stability or change in mental health status already exist. Prior research on, or experience with, patients and their lives can, otherwise, only serve to suggest possibilities to the relevant public.

Whatever is the appropriate content for an evaluative assessment of patients, some adequate means of obtaining the pertinent data must be developed. The criterion data, according to our analysis of disease, are on patient-role performance capacities (which are dispositional properties of persons) and on patient experiences, acts, and events *tending* to alter these capacities (which are state properties of persons with certain dispositional and causal properties of their own).

A person's role-performance capacities are fully manifest by his role performance only when his motivation to perform is optimal and his opportunities to perform or conditions of performance are sufficiently similar to those of interest. Thus, proper test conditions for mental health status require sufficient motivation of the patient for adequate role performance and "real life" situations appropriate to this performance. Only patients themselves, their natural associates, or a very intrusive evaluation team (made more subtle, perhaps, by modern electronic apparatus)

could observe patients under these conditions. Although it may be reasonable to suppose that inadequate role performance in one situation is predictive of similar inadequacy in other situations, whether this is so or not is an empirical matter. Certainly performance in the unfamiliar role of patient, test taker, applicant for help, or prisoner *may* not be a good predictor. Since therapists typically know only what they are told by patients and collaterals and what they see *in* the therapy or testing situation, their assessments are either (adjusted) hearsay evidence or prognostications, some of which are generalizations to other situations. Therefore, evaluative research requires that good working relations be established with patients and their natural associates for interviewing or, better, observation by the evaluation team. The subtle problems of bias, reactivity, and indirect measures of role performance are not taken up here.

As noted in the preceding paragraphs, the presence of any experiences of, acts by, or events in the patient which tend to incapacitate him are also within the evaluative purview. Their disposition to incapacitate is an empirical matter on par with therapists' prognostications, but some tend to be somewhat incapacitating by definition. Thus, intolerable anxiety or strain, serious and apparently insoluble "life problems," unpleasant irrational behavior, and flagrant antisocial behavior are prime examples. Whatever the whole gamut of these "pathogenic" states of the person may be, those which are acts or (nonexperiential) events are accessible to natural associates, evaluation teams, therapists, and some patients if they observe the patient for a sufficient length of time and over a sufficient variety of circumstances. The likely error is that of failure to note transitory relevant behaviors. Although the same is true for relevant patient experiences, only the patient himself has access to these and so for this data his therapist should have a considerable advantage over many of the patient's natural associates and even over many evaluation team interviewers or observers. Again in the case of these several states of the patient, any further validity or reliability issues are not considered.

⁷ An example of an evaluative study conducted as an application of the logic of this essay, and particularly of this and the following section, is Krause, Breedlove, and Gonen (1967).

DECIDING BETWEEN TREATMENTS

Once it is admitted that several legitimate standpoints for evaluation exist and, further, that each involves several criteria, the problem of how to choose among treatments on the basis of their outcomes becomes very complicated. Only when these outcomes are superior for one treatment on all criteria from all standpoints is a choice simple. Obviously, it *can* happen that what is an improvement from one standpoint (e.g., to government authorities a reduction in unconventional behavior) may be taken to be worse from another (e.g., to humanistic therapists a reduction in self-actualization). And evaluative research cannot avoid the implications of such conflicts of values in pluralistic societies. Perhaps the easiest solution derives from using the values of those who sponsor the research (and see some decisions to be made on the basis of its results). Less interested parties' values may be involved if there are sufficient resources to collect the relevant data and if they too can be informed of the results. This seems a more socially responsible solution, but it demands that these various parties be identified and their standpoints defined first.

Even within a single standpoint several noncolinear or even negatively correlated criteria may occur. If each could be assigned some proper weight, then a unique outcome score could be calculated for each case. But it is difficult to see, without empirical study, how important each criterion is in each standpoint or whether the criteria are properly

additive at all. In the meantime, all that is clear is that a case that improves on some criteria and becomes worse on none is better than another case which improves less or gets worse on any of these same criteria and does not improve more on any of them. Where other patterns of differences occur, choices between treatments require prior weighting rules for the criteria.

REFERENCES

- DUBOS, R. *Mirage of health*. New York: Doubleday, 1961.
- JASPERS, K. *General psychopathology*. Manchester: Manchester University Press, 1962.
- KRAUSE, M. S. The construct validity of measuring instruments. *Journal of General Psychology*, 1967, 77, 277-284.
- KRAUSE, M. S., BREEDLOVE, J. L., & GONEN, K. S. An experiment in family therapy for alcoholism. Research report. Available from author, 1967. (Mimeo)
- MARGOLIE, J. *Psychotherapy and morality*. New York: Random House, 1966.
- MASLOW, A. H. *Toward a psychology of being*. Princeton: Van Nostrand, 1962.
- MECHANIC, D. Some factors in identifying and defining mental illness. *Mental Hygiene*, 1962, 45, 66-74.
- OFFER, D., & SABSHIN, M. *Normality*. New York: Basic Books, 1966.
- PARSONS, T. *Social structure and personality*. New York: Free Press, 1964.
- SZASZ, T. S. *The myth of mental illness*. New York: Hoeber-Harper, 1961.
- TEMKIN, O. The scientific approach to disease: Specific entity and individual sickness. In A. C. Crombie (Ed.), *Scientific change*. London: Heinemann, 1963.
- WOOTTON, B. *Social science and social pathology*. New York: Macmillan, 1959.

(Received July 10, 1968)

CROSS-CULTURAL COMPARISON OF PSYCHOTIC SYNDROMES¹

MAURICE LORR²

Catholic University

AND

C. JAMES KLETT

*Veterans' Administration Hospital,
Perry Point, Maryland*

The equivalence of 12 psychotic syndromes established in the United States data was sought in data from six other countries. Samples of drug-free, newly hospitalized patients in England, France, Germany, Italy, Japan, and Sweden were rated after a semistructured interview. The pooled ratings of men and women in each sample were correlated and factored. A hypothesis-testing procedure and the normal Varimax were applied to yield two roughly equivalent solutions. With but a few exceptions each of the syndromes was confirmed in each of the six samples. The sample syndrome score profiles were also examined and discussed.

Beginning in 1933 with Moore, a group of investigators have sought to isolate, by the method of factor analysis, the major psychotic syndromes reflected in currently observable symptoms and behavior. The need arose from the prevailing dissatisfaction with the psychiatric classes and from the absence of any rigorous alternative procedure for correcting or verifying the symptom groupings reported by clinicians. Wittenborn (1951), Guertin (1952), Lorr, Jenkins, and O'Connor (1955), and Lorr, McNair, Klett, and Lasky (1962) have among them isolated and confirmed 10 or more psychotic syndromes. Since a critical review of these and other studies has been reported (Lorr, Klett, & McNair, 1962), none is needed here.

The present study proposed to ascertain whether or not the psychotic syndromes defined by the factor analysis of interview ratings were the same in other cultures than our own. The major goal was to determine the number and kinds of syndromes exhibited by drug-free and newly hospitalized men and women in each of six countries. The results were then compared across samples and with

the United States findings to assess the degree of constancy or similarity. At the present time there is no internationally accepted yardstick for measuring the kinds or degrees of psychotic behavior manifested by the severely disturbed. Qualitative descriptions and categories constitute the major vehicles for communicating research information concerning the behavioral disorders. In brief, the investigation could provide a test of the uniformity of the syndromes across several cultures. The end results would perhaps yield a set of measures and a conceptual scheme capable of facilitating epidemiological, genetic, and biochemical research.

The 12 syndromes or constructs hypothesized were as follows: Excitement, Hostile Belligerence, Paranoid Projection, Grandiosity, Perceptual Distortion, Obsessional-Phobic, Anxious Depression, Functional Impairment, Retardation and Apathy, Disorientation, Motor Disturbances, and Conceptual Disorganization. A Somatic Complaints syndrome was also framed but dropped for reasons to be given later. Each of these patterns has been identified two or more times. They also combine into five more inclusive higher-order syndromes (Lorr, Klett, & Cave, 1967) of which four resemble the well-known disorders referred to as schizophrenia, paranoid state, psychotic depression, and manic excitement. At the same time it should be understood that the syndromes hypothesized, based as they are on current symptoms and signs, are not synonymous with the usual diagnostic

¹ This study received partial support from National Institutes of Health Grant MH 113212-02. Grateful acknowledgment is made to Horst Flegel, Franco Giberti, Max Hamilton, Carl Jonsson, Ziro Kaneko, and Pierre Pichot who made this study possible. Ronald Bonato and Kenneth Yang of the Biometrics Laboratory, George Washington University, conducted the initial factor analyses.

² Requests for reprints should be sent to Maurice Lorr, Department of Psychology, Catholic University of America, Washington, D. C. 20017.

classes. Conventional psychiatric categories include additionally such life history variables as age, type of onset, precipitating event, course of the disorder, and tendency toward recurrence.

METHOD

Data Collection

Conferences were held with each group of collaborating psychiatrists in England, France, Germany, Italy, and Sweden. Arrangements with the Japanese representative, however, were conducted by mail. All psychiatrists responsible for data collection held university appointments. The meetings were concerned with the nature of the sample, with the interview structure, and with the rating process. Several patients were interviewed and rated in order to familiarize each team with the rating schedule and with the rating process. Subsequently arrangements were made to translate the rating form of the Inpatient Multidimensional Psychiatric Scale (IMPS) into French, German, Italian, Swedish, and Japanese. The forms were then retranslated back into English by another competent person. Finally the original was compared with the retranslation by the investigators.

Each team was asked to interview and rate a representative sample of 100 males and 100 females diagnosed psychotic and newly hospitalized. A further requirement was that patients interviewed either be free of drugs or be on a minimum drug-dosage schedule. This requirement arose from research showing that the tranquilizers reduced or eliminated characteristic symptom profiles. Cases with central nervous system disorders or neurological disease as well as alcoholics and drug addicts were to be excluded.

In addition to the ratings, data were collected on the following: age, marital status, highest grade completed, occupation, place of residence (rural or urban), initial psychiatric diagnosis, religion, duration of psychotic episode, and number of prior hospitalizations.

Interview and Rating Schedule

The standardized 75-item rating schedule IMPS was designed to measure 11 psychotic syndromes isolated in repeated factor analyses (Lorr, Klett, McNair, & Lasky, 1962). To the standard set, 15 new items were added to measure two depression syndromes recently established called Functional Impairment and Somatic Complaints (Lorr, Sonn, & Katz, 1967). Five to 11 items defined each syndrome measure. Data relative to the reliability and validity of the standardized portion of IMPS are presented elsewhere (Lorr & Klett, 1966; Lorr, Klett, McNair, & Lasky, 1962).

The interviews were semistructured, 30 to 60 min. in length, and focused on information relevant to the questions in IMPS. Each interviewer was accompanied by an equally experienced and trained observer. Ratings were completed independently by observer and interviewer immediately following the

interview. It was the combined ratings that were used in the subsequent analysis. The interviewing team also determined the patient's diagnosis using the American Psychiatric Association (1952) manual.

Method of Analysis

The first step was to ascertain the number of factors required to account for the correlations among the items. The two criteria used to judge completeness of factoring were (a) the number of factors with associated latent roots of 1.00 or greater, and (b) the number of Varimax-rotated factors, exclusive of doublets and singlets, judged to be meaningful. To this end two preliminary principal component analyses were made, one of the pooled male cases and one of the pooled female cases. The two 14-factor solutions yielded 12 interpretable Varimax factors, of which 11 were clearly the same in both analyses. The analysis also revealed that the new Somatic Complaints syndrome hypothesized was represented weakly or was absent in several countries. For this reason the five somatic items were deleted for the hypothesis-testing portion of the study which was based on 85 variables.

Next each sample's correlation matrix, based on the pooled ratings of both men and women, was factored by the method of principal components. Pooling seemed justified by prior research indicating negligible sex differences (Lorr & Klett, 1965), and by the reduction in the standard errors in correlation coefficients. The 12 factors extracted in each sample were rotated by a least-squares procedure to a simple structure hypothesis matrix (Horst, 1956) based on United States data. The simple structure matrix consisted of 85 rows and 12 columns representing the items and the factors hypothesized. Each item was assigned a zero weight except in the factor column it served to define. The nonzero weights ranged .45-.85.

The third set of analyses were based on all 90 variables. Thirteen principal components were extracted from each of the six matrices. The factors obtained were then transformed by Kaiser's normal Varimax. The use of 13 rather than 12 factors (as in the second analysis) was justified by the more stable structure attained in rotation. The "blind" analytical solution was intended as a check on the hypothesis-testing solution. The latter should provide a better fit to the data (being a correlated solution) and at the same time test the equivalence of United States and cross-cultural structures.

RESULTS

Sample Characteristics

The total sample consisted of 1,107 cases to which each country except Sweden contributed between 200 and 204 cases. The Swedish sample consisted of 50 men and 50 women. The median age of the 540 men was 36 while the median age of the 567 women

TABLE 1
CORRELATIONS OF SYMPTOMS WITH EXCITEMENT SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Unrestrained behavior	7	.59	.71	.64	.44	.49	.53
Hurried speech	9	.69	.71	.68	.74	.67	.63
High mood	12	.51	.61	.42	.41	.62	.45
Dramatizes	17	.54	.45	.58	(.33)	.66	.64
Loud speech	20	.65	.62	.65	.71	.66	.64
Overactive	26	.68	.76	.62	.59	.65	.61
Excess speech	35	.72	.70	.60	.78	.66	.66
Dominating	37	.64	.59	.47	.72	.60	.65
Intense feeling	76	.51	.36	.49	.76	.44	.50
Seductive	77	.50	—	.49	—	.57	.45
	Not keyed						
Rambling speech	4	.44	.51	—	.42	.51	.43
Superior attitude	15				.35	.51	.36
Tension	10			.37	.35		
Irritability	25			.42			
Slovenliness	30		.36				
Complaining	34				.39		
Sleep difficulty	86		.37				

Note.—Number in parentheses not significant at $\pm .35$.

was 38. The average years of schooling received by patients ranged from 6.5 for Italians to 11.5 for the Japanese. The percentage of single (never married) was 63 for men and 37 for women. No previous hospitalization had been experienced by 46% of the men and 43% of the women. As might be expected nearly all of the French and Italian patients

were Catholic and nearly all of the Swedish were Protestants. The English and the German sample was split with Protestants predominating. Two-thirds of the Japanese were Buddhists. Approximately half of the sample came from very large cities. The occupation of most patients fell into clerical and sales, skilled and semiskilled.

TABLE 2
CORRELATIONS OF SYMPTOMS WITH ANXIOUS DEPRESSION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Blames self	14	.79	.79	.72	.78	.69	.79
Anxious (specific)	21	.45	.50	.54	.46	.47	.57
Apprehensive	24	.44	.44	.51	.42	.41	.47
Self-depreciating	27	.71	.73	.63	.68	.62	.74
Depressed	29	.45	.54	.53	.40	.59	.65
Guilty	31	.83	.83	.77	.78	.71	.77
Lacks insight	—39	(— .26)	—	—	(— .34)	—	—
Suicidal	40	.44	.50	.54	.35	.35	.41
Feels sinful	66	.72	.77	.64	.61	.57	.68
Hopeless	78	.56	.60	.56	.42	.65	.63
	Not keyed						
Tense	10	.43				.39	
Slow moving	16						
Obsessive	41	.37					
Lacks sex interest	81						.38

Note.—Numbers in parentheses not significant at $\pm .35$.

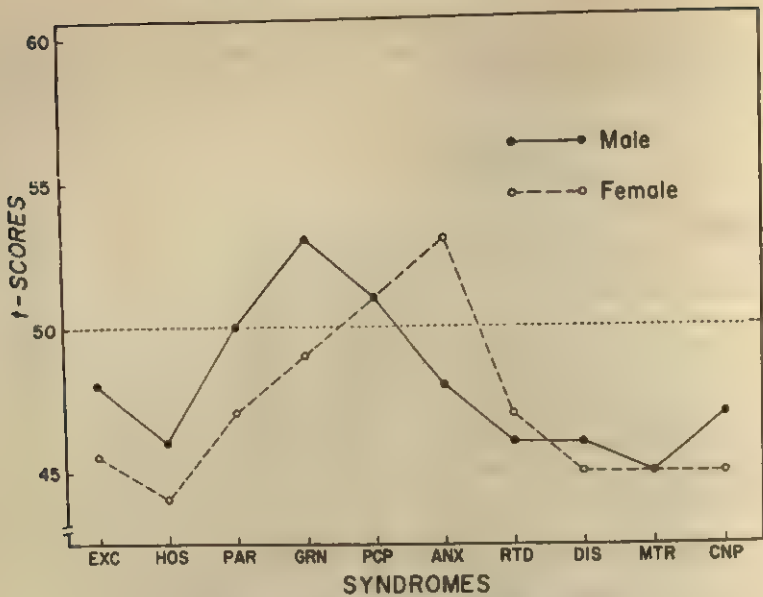


FIG. 1. Mean syndrome profiles for English sample.

A standard American Psychiatric Association (1952) diagnosis was obtained for each case. The English, French, German, and Swedish samples were fairly consistent in their distribution of cases to the various diagnostic categories. The percentages of nonparanoid schizophrenics, paranoids (including paranoid schizophrenics), depressives, and manics were 40, 30, 22, and 8, respectively. In contrast, more than half of the Italian sample were diagnosed depressives and about half of the Japanese received nonparanoid schizophrenic diagnoses.

Hypothesis-Testing Solution

Tables 1 to 12 present by country those symptoms and behaviors correlating $\pm .35$ with each syndrome in the hypothesis-testing solution. The items keyed for the syndromes (those given nonzero weights) are shown in the upper section of each table. The items not keyed but correlating above .35 with the factor are shown in the lower section of each table. Items 76 to 90 represent the experimental scales.

The results are summarized briefly as the abbreviated symptom descriptions make it

TABLE 3
CORRELATIONS OF SYMPTOMS WITH FUNCTIONAL IMPAIRMENT SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Lacks sex interest	81	(.26)	.50	—	(.32)	.36	.36
Cannot work	84	.49	.57	.59	.70	.48	.53
Cannot concentrate	85	.46	.59	.59	.62	.42	.49
Sleep difficulty	86	.41	.41	.56	.50	.49	.39
Fatigued	87	.54	.59	.50	.61	.51	.40
No interest in people	90	.63	.53	.37	.64	.44	.52
	Not keyed						
Apprehensive	24	.39					
Lacks insight	39						— .49
Suicidal	40					.35	
Hopeless	78	.37					

Note.—Numbers in parentheses not significant at $\pm .35$.

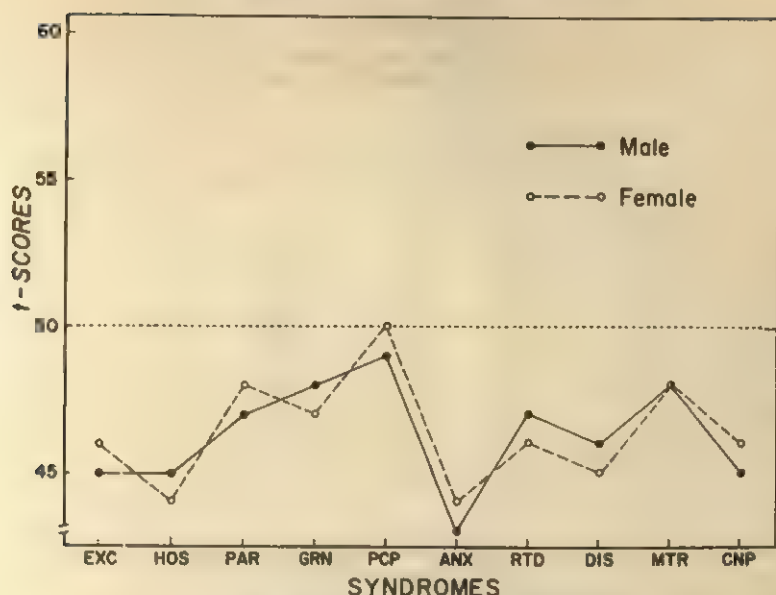


FIG. 2. Mean syndrome profiles for French sample.

obvious what defines a factor. Nonkeyed items are considered if they correlated with a factor in at least three countries.

Excitement. This pattern presented in Table 1 emerged similarly in each country. Rambling speech (4) and an attitude of superiority (15), although not in the key, also correlated with the syndrome. In the United States samples they also correlated with Excitement but at lower levels.

Anxious Depression. The pattern presented in Table 2 corresponds closely to the United States key except for lack of insight which was expected to correlate negatively.

Functional Impairment. This less well-established syndrome presented in Table 3 was

defined fairly uniformly in each of the six samples as hypothesized. It had been previously identified by Overall (1963) and by Lorr, Sonn, and Katz (1967).

Obsessional-Phobic. Table 4 presents a pattern fairly well defined in four countries, weakly defined for France, and absent from Sweden. Ideas of change or depersonalization (67) was also weak in the United States data as it is here.

Retardation and Apathy. With but a few exceptions this syndrome (Table 5) was confirmed in each sample. Manneristic posturing (6) was the only nonkeyed item relating to Retardation in three samples.

TABLE 4
CORRELATIONS OF SYMPTOMS WITH OBSESSIVE-PHOBIC SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
Obsessive Fearful (specific) Compulsive Ideas of change	Keyed						
	41	.38	(.26)	.62	.79	.53	—
	42	.50	(.31)	.38	.75	.65	—
	43	—	—	.57	.80	.37	—
Neologisms External control	67	.51	.43	—	—	—	—
	Not keyed						
	49		.37				
	63		.38				

Note.—Numbers in parentheses not significant at $\pm .35$.

TABLE 5
CORRELATIONS OF SYMPTOMS WITH RETARDATION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Slowed speech	1	.68	.57	.51	.72	.59	.61
Lacks goals	8	.53	.47	.55	(.32)	.66	.59
Fixed facies	13	.69	.68	.63	.62	.67	.65
Slowed movements	16	.74	.65	.48	.72	.54	.56
Memory deficit	19	(.32)	.39	(.29)	.49	.45	—
Blocked speech	22	.51	.44	(.34)	—	.47	.48
Apathetic	23	.57	.54	.59	.47	.71	.59
Slovenly	30	.42	.59	(.33)	.47	(.30)	.38
Whispered speech	33	.67	.68	.50	.72	.57	.52
Fails to answer	38	.61	.67	.67	.71	.61	.61
	Not keyed						
Postures	6	.36	.43			.51	
Irrelevant speech	2				.41		
Lacks insight	39					.61	
Repet. movements	48					.48	

Note.—Numbers in parentheses not significant at $\pm .35$.

Hostile Belligerence. The belligerence pattern shown in Table 6 was clearly confirmed in each sample. Ideas of persecution (60) was the only nonkeyed item represented.

Disorientation. Although relatively few patients were rated disoriented, the defining items (Table 7) were relatively objective and reliable. This may be the reason the factor appeared uniformly in each sample.

Grandiosity. The pattern of grandiose expansiveness shown in Table 8 was relatively narrow but was well established. Here it was most weakly represented in Italy.

Conceptual Disorganization. The items that define the pattern shown in Table 9 are associated with process-type disorganized schizophrenia. Thus it is understandable why the syndrome fails to appear in Italy and in

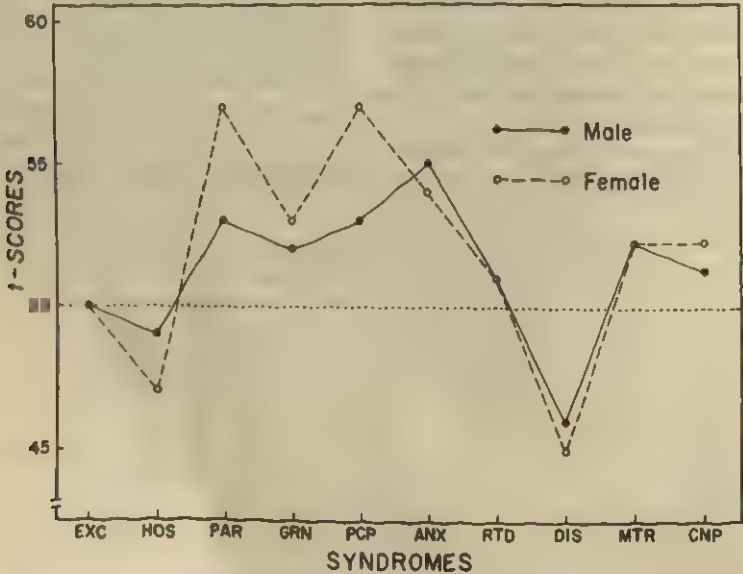


FIG. 3. Mean syndrome profiles for German sample.

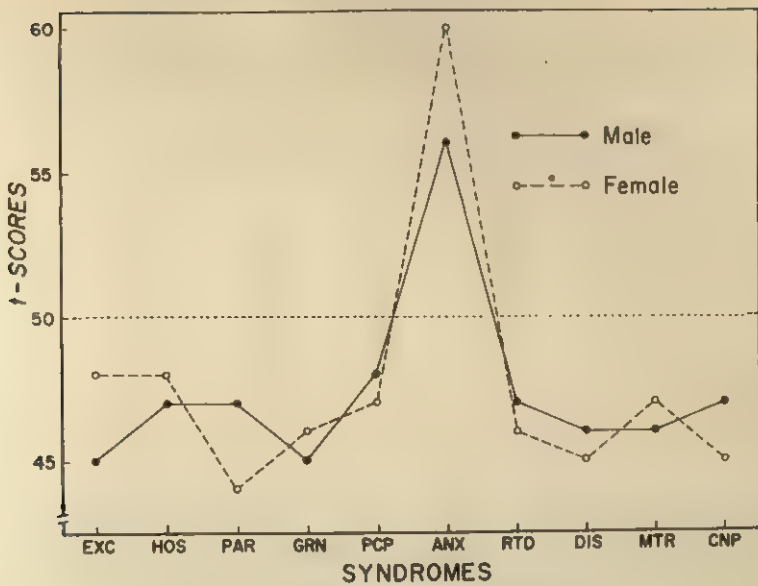


FIG. 4. Mean syndrome profiles for Italian sample.

Sweden. The Italian sample consisted mainly of depressives while the Swedish sample was restricted to mild cases by the requirement that patients be drug free.

Motor Disturbances. The syndrome, shown in Table 10, was confirmed in four samples, weak in Sweden, and absent from Japan. Since nonparanoid diagnoses were well represented in the Japanese sample it was difficult to account for the absence of the syndrome. Speech blocking (Item 22) correlated better with

Motor Disturbances than with Conceptual Disorganization where it was keyed.

Perceptual Distortion. The syndrome shown in Table 11 is characterized by voices, visions, and other hallucinatory experiences. Since it was difficult to eliminate content, it is not surprising that Item 54 (voices that extol) correlated with the syndrome and with Grandiosity.

Paranoid Projection. Table 12 shows that the paranoid syndrome was well characterized

TABLE 6
CORRELATIONS OF SYMPTOMS WITH HOSTILE BELLIGERENCE SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Verbally hostile	5	.73	.76	.72	.67	.73	.77
Contemptuous	11	.42	.69	.60	.60	.45	.52
Hostile attitude	18	.71	.77	.70	.50	.65	.64
Irritable	25	.64	.71	.53	.49	.54	.64
Blames others	28	.58	.51	.61	.62	.67	.70
Resentful	32	.63	.61	.69	.63	.75	.78
Complaining	34	.68	.73	.68	.51	.57	.69
Suspicious	36	.44	.69	.54	.45	.51	.66
	Not keyed						
Ideas of persecution	60	.47	.36	.43			
Tense	10		.46				
Dramatizing	17			.40			
Slovenly	30						.40
Lacks insight	39						.35

TABLE 7
CORRELATIONS OF SYMPTOMS WITH DISORIENTATION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Knows hospital	70	.57	.69	.93	.91	—	.60
Knows state	71	.49	.72	.83	.94	.43	.61
Knows no one	72	.55	.68	.62	.76	.44	—
Knows season	73	.64	.76	.90	.94	.59	.78
Knows year	74	.66	.77	.86	.81	.62	.80
Knows age	75	.45	.80	.67	.80	.62	.75
	Not keyed						
Memory deficit	19		— .35				
Stereotyped speech	50						— .42
Ideas of change	67						.35

in each sample. Item 36, suspicion, although keyed for Hostile Belligerence also correlated moderately with the pattern. In the English, French, and German samples, four of the variables keyed for Perceptual Distortion also correlated with the paranoid syndrome. Several hypotheses can be offered to explain the extent of overlap between the two syndromes. First, it may be conjectured that the interviewers in some countries are more skillful or more thorough in seeking out hallucinations when paranoid ideation is present. Or, it is possible that because of tradition, raters in

some countries tend to assume the presence of hallucinations when paranoid delusions are noted. A third hypothesis is that the hallucinating experiences are more frequently associated with persecutory ideas in some countries than others. Further research with structured interviews may clarify the problem.

The correlations among the primary factors of each solution were next examined. If these are quite large, they would suggest that the solution had been forced. However, the correlations were negligible or small with but few exceptions. Moderate correlations occurred at

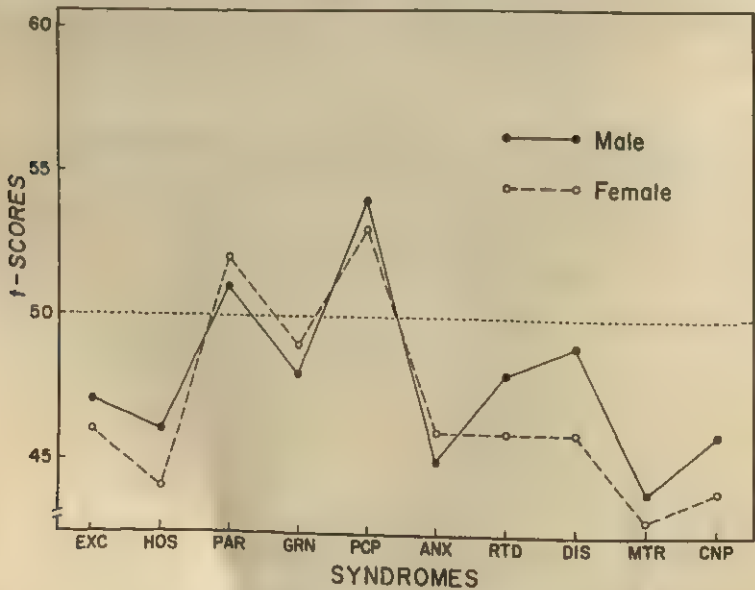


FIG. 5. Mean syndrome profiles for Japanese sample.

TABLE 8
CORRELATIONS OF SYMPTOMS WITH GRANDIOSITY SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Superiority	15	.50	.40	.41	.48	(.27)	.35
Voices extol	54	.68	.59	.50	—	.43	.51
Unusual powers	64	.56	.66	.61	.54	.68	.66
Great personality	65	.66	.62	.63	.72	.74	.39
Divine mission	69	.67	.68	.74	.73	.63	.60
	Not keyed						
Visions	57		.40	.35			
Elevated mood	12	.37					
Specific fears	42						— .38
Voices order	56	.53					
Amorous delusions	79						.57

Note.—Numbers in parentheses not significant at $\pm .35$.

TABLE 9
CORRELATIONS OF SYMPTOMS WITH CONCEPTUAL DISORGANIZATION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Irrelevant speech	2	.56	.41	.56	—	.67	—
Incoherent speech	3	.62	.49	.67	—	.64	—
Rambling speech	4	.45	.37	.41	—	.48	—
Neologisms	49	.49	.58	.35	—	.52	—
Stereotyped speech	50	.45	—	.45	—	(.34)	—
	Not keyed						
Startled glances	52					.36	

Note.—Numbers in parentheses not significant at $\pm .35$.

TABLE 10
CORRELATIONS OF SYMPTOMS WITH MOTOR DISTURBANCES SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Posturing	6	.57	.44	.64	.59	—	.72
Tense	10	(.27)	—	(.29)	(.30)	—	.39
Giggling	46	.48	.62	.50	.64	—	—
Grimacing	47	.74	.57	.66	.64	—	.71
Repet. movements	48	.73	.53	.67	.71	—	.71
Talking to self	51	.63	.47	—	.39	—	(.33)
Startled glances	52	.47	.46	.41	.42	—	.35
	Not keyed						
Speech blocking	22	.44	.45		.38		.36
Irrel. speech	2		.44				
Incoherent speech	3				.48		
Slovenly	30						.36
Stereo. speech	50				.54		
Seductive behavior	77		.51				

Note.—Numbers in parentheses not significant at $\pm .35$.

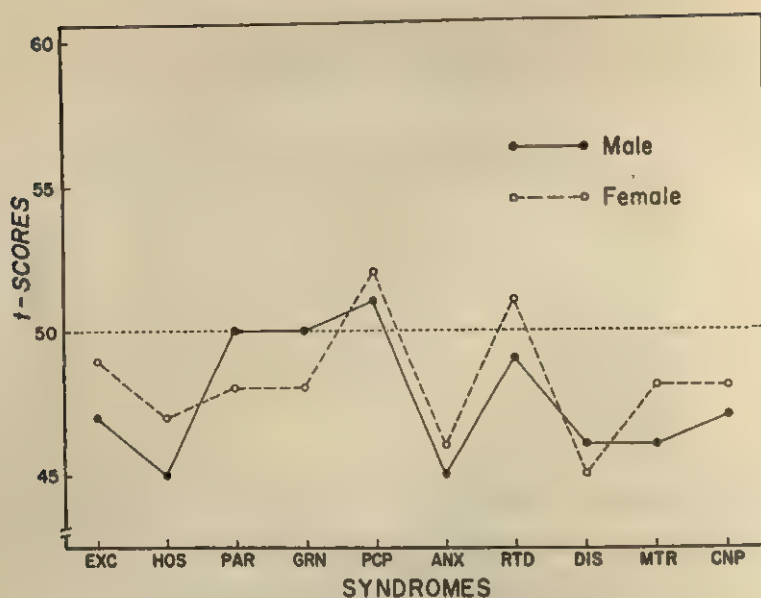


FIG. 6. Mean syndrome profiles for Swedish sample.

approximately the same level obtained in the United States data.

Varimax Solution

The six Varimax solutions were examined with a view to finding matches across samples and with the United States key. The cutting point for a significant correlation was arbitrarily set at $\pm .35$, a widely accepted value. On the whole, the simple structures achieved

by the Varimax solution were not as clear as in the main solution. Although the syndromes tended to be correlated, the Varimax yielded independent factors. Consequently, items were likely to correlate significantly with several factors rather than one. However, six factors did have clean matches across the various samples. These factors were Excitement, Perceptual Distortion, Grandiosity, Anxious Depression, Retardation, and Disorientation.

TABLE 11
CORRELATIONS OF SYMPTOMS WITH PERCEPTUAL DISTORTION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
Hears voices	Keyed 45	.75	.75	.68	.59	.73	.62
Voices accuse	53	.59	.66	.60	.72	.60	.53
Voices threaten	55	.49	.60	.63	.54	.63	.69
Voices order	56	.55	.56	.69	.53	.62	.36
Visions	57	.52	—	.45	.42	.41	.35
Hallucinations (other)	58	.44	.39	.49	.37	.37	.63
Voices extol	Not keyed 54	.37	.38	.42	.40	.44	
Delusions	44	.44	.38				
Startled glances	52			.39	.49		
People control	62	.50	.43				
External control	63	.44					.45
Ideas of reference	59		.43				
Ideas of conspiracy	61	.36					

TABLE 12
CORRELATIONS OF SYMPTOMS WITH PARANOID PROJECTION SYNDROME

Symptom	Item	Country					
		England	France	Germany	Italy	Japan	Sweden
	Keyed						
Delusions	44	.54	.61	.65	.62	.45	.56
Ideas of reference	59	.48	.70	.74	.59	.56	.59
Ideas of persecution	60	.50	.43	.56	.64	.63	.51
Ideas of conspiracy	61	.63	.34	.59	.61	.59	.55
People control	62	.51	.57	.64	.63	.43	.39
External control	63	.38	.36	.49	.53	.53	(.27)
Amorous delusions	79	.34	—	.37	—	(.26)	.50
Homosexual delusions	80	.42	.34	.36	—	(.29)	.44
	Not keyed						
Suspensions	36	.51	.35	.61	.56		.43
Hears voices	45	.39	.53	.42			
Voices accuse	53	.50	.51	.43			
Resentment	32			.35	.43		
Voices threaten	55	.39	.36				
Hallucinations (other)	58	.40		.40			
Ideas of change	67		.49	.60			
Verbal hostility	5			.36			
Tensions	10					.36	
Blames others	28				.35		
Voices order	56		.36				
Visions	57	.43					
Great personality	65						.41

Note.—Numbers in parentheses not significant at $\pm .35$.

Four additional factors were represented in five countries. These factors were Hostile Belligerence, Paranoid Projection, Functional Impairment, and Motor Disturbance. Obsessional-Phobic and Conceptual Disorganization patterns were represented in four samples. The Somatic Disturbance syndrome appeared in three countries. It could not have appeared in the hypothesis-testing analyses as the five defining items had been deleted.

The Varimax findings may also be viewed

from the standpoint of each country. All syndromes except Functional Impairment and Obsessional-Phobic appeared in the English data. The French sample yielded no Paranoid Projection syndrome; paranoid items correlated instead with Hostile Belligerence. The German sample failed to produce a Hostile Belligerence or a Somatic Disturbance syndrome. The syndromes called Conceptual Disorganization and Somatic Disturbance were missing from the Italian sample. The Japa-

TABLE 13
SYNDROME RAW SCORES FOR MALES

Syndrome	Country						
	England	France	Germany	Italy	Japan	Sweden	United States
Excitement	18	10	23	10	12	15	20
Hostile belligerence	20	18	30	22	15	18	31
Paranoid projection	35	26	41	24	39	34	33
Grandiosity	14	6	13	3	8	9	9
Perceptual distortion	15	12	18	9	20	15	13
Anxious depression	34	20	53	58	28	25	38
Retardation	26	28	40	28	25	34	37
Disorientation	0	0	0	0	1	1	1
Motor disturbances	11	15	21	12	6	11	18
Conceptual disorganization	8	5	14	8	4	8	11

TABLE 14
SYNDROME RAW SCORES FOR FEMALES

Syndrome	Country						
	England	France	Germany	Italy	Japan	Sweden	United States
Excitement	16	14	27	21	18	24	25
Hostile belligerence	18	18	28	31	25	27	37
Paranoid projection	27	31	55	19	38	31	34
Grandiosity	8	5	12	3	7	5	8
Perceptual distortion	16	15	28	9	20	17	14
Anxious depression	49	23	51	68	26	31	40
Retardation	25	23	38	25	30	38	34
Disorientation	0	0	0	0	1	1	1
Motor disturbances	9	15	21	13	8	15	18
Conceptual disorganization	6	7	14	6	7	9	12

nese sample produced no Motor Disturbance factor while the Swedish sample showed little evidence of Obsessional-Phobic, Somatic Disturbance, or Conceptual Disorganization. No tables are presented because the results are quite similar to those presented earlier.

The findings from the two solutions thus support the conclusion that the same syndromes exist in all seven cultures. However, there were minor differences between countries as to the variables that characterize each syndrome. It seems likely that these differences reflect mainly sample biases.

Culture Syndrome Differences

Since the syndromes were comparatively constant across the seven cultures it was possible to examine between-country differences in score profile. Each case was scored on the 10 standard syndromes on the basis of the United States derived keys and weights. The mean syndrome scores were computed for both men and women by country. Table 13 illustrates the range of variation in men. The mean raw scores for each country were also converted into standard scores (based on United States norms) in order to examine the profiles.

When the samples were ranked with respect to mean score on each of the syndromes it became apparent that the German sample was the most severely disturbed and the United States sample next. The French and Swedish samples were least disturbed; mean profile elements were nearly all below the American norms.

It was also possible to characterize the mean syndrome standard score profile of each sample in terms of one or two scores most elevated above the norms. The English males sample was then seen as primarily grandiose and the female sample as anxious depressed. While the French profiles were relatively flat they exhibited peaks on paranoid projection and hallucinations. The German males may best be characterized as anxious depressed and paranoid. The female German sample tended to be hallucinated paranoids. Both Italian samples peaked only on anxious-depression, a finding consistent with diagnoses previously described. The Japanese samples were also characterized by elevated scores on paranoid projection and perceptual distortion (hallucinations). The Swedish sample could not be characterized in any particular way since the profile was comparatively flat.

DISCUSSION

The findings from the structural analyses are relatively clear-cut. The same set of syndromes appear consistently across all samples with only minor differences. If somewhat broader samples could have been obtained it seems likely that every syndrome would have appeared in each sample.

The question may be raised as to whether and to what extent the findings reflect professional training and knowledge of Kraepelinian categories. It should be noted in response, that at least half of the syndromes are not recognized as such by psychiatrists. Hostile Belligerence, Grandiosity, Perceptual

Distortion, and Conceptual Disorganization are examples. A second and equally cogent answer is that the same syndromes have been isolated in ratings provided by psychiatric aides on ward behavior, and by patient relatives on patient home behavior.

The differences between countries in mean syndrome scores should be considered, at best, as suggestive. It is for this reason that no significance tests were made. The samples simply are not representative of any country's hospitalized psychotic population; this was not a study goal. Although representative samples from each hospital's intake were sought, a greater emphasis was placed on the inclusion of all sources of individual variation in psychopathology. A sample designed for a factor analysis need not be representative of a population but it should include the range of variation on all major sources of difference. For example, to define a height factor it is necessary to include tall and short people; a random sample is not required. Repeated factor analysis of widely different United States samples supports this argument. Studies have been made of state hospital cases, veterans hospital cases, and university-sponsored hospital cases. All yielded the same structures provided the samples were stratified on the spectrum of psychotic behavior.

REFERENCES

AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and statistical manual for mental disorders*. Washington, D. C.: Mental Hospital Service, 1952.

- GUERTIN, W. H. A factor-analytic study of schizophrenic symptoms. *Journal of Consulting Psychology*, 1952, 16, 308-312.
- HORST, P. A simple method of rotating a centroid factor matrix to a simple structure hypothesis. *Journal of Experimental Education*, 1956, 24, 231-237.
- LORR, M., JENKINS, R. L., & O'CONNOR, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *Journal of Abnormal and Social Psychology*, 1955, 50, 78-86.
- LORR, M., & KLETT, C. J. Constancy of psychotic syndromes in men and women. *Journal of Consulting Psychology*, 1965, 29, 309-313.
- LORR, M., & KLETT, C. J. *Inpatient Multidimensional Psychiatric Scale*. Manual. Palo Alto, Calif.: Consulting Psychologists Press, 1966.
- LORR, M., KLETT, C. J., & CAVE, R. Higher level psychotic syndromes. *Journal of Abnormal Psychology*, 1967, 72, 74-77.
- LORR, M., KLETT, C. J., & MCNAIR, D. M. *Syndromes of psychosis*. Oxford: Pergamon Press, 1962.
- LORR, M., KLETT, C. J., MCNAIR, D. M., & LASKY, J. J. *Inpatient Multidimensional Psychiatric Scale*. Manual. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- LORR, M., MCNAIR, D. M., KLETT, C. J., & LASKY, J. J. Evidence of ten psychotic syndromes. *Journal of Consulting Psychology*, 1962, 26, 185-189.
- LORR, M., SONE, T. M., & KATZ, M. M. Toward a definition of depression. *Archives of General Psychiatry*, 1967, 17, 183-186.
- MOORE, T. V. The essential psychoses and their fundamental syndromes. *Studies in Psychology and Psychiatry*, 1933, 3, 1-128.
- OVERALL, J. E. Dimensions of manifest depression. *Journal of Psychiatric Research*, 1963, 1, 239-245.
- WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *Journal of Consulting Psychology*, 1951, 15, 290-302.

(Received July 12, 1968)

COGNITIVE FACTORS IN DIFFERENTIAL CONDITIONING OF THE GSR:

USE OF A REACTION TIME TASK AS THE UCS WITH NORMALS AND SCHIZOPHRENICS¹

PAUL E. BAER² AND MARCUS J. FUHRER³

Baylor University College of Medicine

Twenty-two normals and 22 schizophrenics underwent differential conditioning of the GSR, using 8-sec. tones of differing frequencies as the CSs, an 8-sec. CS-UCS interval, and a UCS comprised of an RT task signaled by a low-intensity light. Both intertrial reports and postconditioning interviews were obtained. Twelve Ss in the normal group verbalized the conditional stimulus relations accurately, compared to 3 Ss in the schizophrenic group. The normal group showed significant GSR differentiation, though conditioning was limited to the group of accurately verbalizing Ss. No evidence for conditioning was obtained in the schizophrenic group. Normal Ss had faster RTs than schizophrenic Ss. The RT of accurately verbalizing normals was shorter than that of inaccurately verbalizing normals, and the degree of GSR differentiation was significantly correlated with RT.

In recent investigations of cognitive processes during differential classical conditioning, conditional galvanic skin response (GSR) differentiation was found to be limited to groups of Ss capable of accurately describing the conditional stimulus relations during either intertrial reports or postconditioning interviews (Baer & Fuhrer, 1968; Fuhrer & Baer, 1965, 1966). Groups of Ss unable to state the stimulus relations accurately did not show differential GSR conditioning. The Ss in these studies were intellectually capable individuals without known psychopathology. Conditioning

was of an aversive type characterized by the use of a noxious unconditioned stimulus (UCS) consisting of shock at S's tolerance level. Thus, the observed concordance between cognitive and GSR differentiation appeared to represent a relationship between verbalized anticipation of impending shock and GSR indexes of conditional fear responses.

One purpose of this investigation was to extend the study of cognitive processes in differential GSR conditioning by using a non-aversive UCS. The reaction time (RT) signal and its associated motor act appeared promising for this purpose. For instance, Marquis and Porter (1939) have demonstrated that an RT task can serve as an effective UCS for conditioning the eye-blink response. More recently, Martin (1965) showed that the GSR can be differentially conditioned using a UCS comprised of an RT task and a potentially noxious 100-db. tone as the RT signal.

The present study represented an effort to demonstrate that an RT task that is not supplemented by a noxious stimulus can serve as an effective UCS for differentially conditioning the GSR. The use of a long (8-sec.) interstimulus interval permitted the analysis of multiple conditional responses defined by the latency of the GSRs. It was also determined whether there was a relationship be-

¹ Portions of this study were presented to the Society for Psychophysiological Research, San Diego, October 19, 1967. This study was supported in part by United States Public Health Service Grant MH 12908, the Social and Rehabilitation Service, Department of Health, Education, and Welfare, Research and Training Center RT-4, and the Texas Research Institute for Mental Sciences to which the authors are indebted for the patient sample and the biochemical assays. Computer utilization was provided by the Computer Science Program of Baylor University College of Medicine, and the Common Research Computer Facility, supported by United States Public Health Service Grant FR 00254.

² Formerly at Texas Research Institute of Mental Sciences. Requests for reprints should be sent to Paul E. Baer, Department of Psychiatry, Baylor University College of Medicine, Texas Medical Center, Houston, Texas, 77025.

³ Also at Texas Institute for Research and Rehabilitation.

tween cognitive processes and conditional GSR differentiation similar to that found when an aversive UCS is used. In addition, RT was assessed in relation to conditional GSR differentiation and Ss' ability to verbalize stimulus relations accurately.

This study also sought to establish whether a group of schizophrenics (Group Z) would differ from a group of normals (Group N) in conditional GSR differentiation and in the ability to verbalize stimulus relations accurately. There are a number of previously established characteristics of schizophrenics that supported the expectation that Ss in Group Z would be less able to recognize the conditioned stimulus (CS)—unconditioned stimulus (UCS) relations. For example, schizophrenics have been reported to have difficulties in excluding irrelevant stimuli (Lang & Buss, 1965), integrating related stimuli (Chapman & McGhie, 1962), and maintaining an attention set (Zahn, Rosenthal, & Shakow, 1963). It was therefore predicted that there would be fewer accurately verbalizing Ss in Group Z than in Group N, and that there would be less GSR differentiation in Group Z. The latter prediction was based on the assumption that differential GSR conditioning and the ability to verbalize stimulus relations are related for schizophrenics in a manner similar to that previously observed for normal Ss.

METHOD

Subjects and Apparatus

Group N was comprised of 22 women, paid volunteers, whose ages ranged between 19 and 48 ($\bar{X} = 30.8$), and whose educational attainment ranged between 9 and 16 yr. ($\bar{X} = 12.7$). Group Z was comprised of 22 women whose ages ranged between 24 and 47 ($\bar{X} = 34.1$), and whose educational attainment ranged between 8 and 16 yr. ($\bar{X} = 12.5$). The Ss in Group Z were all posthospitalized outpatients who had been prescribed maintenance dosages of phenothiazines. The approximate modal dosage was 100 mg. daily of chlorpromazine. A subgroup of 12 Ss in Group Z were administered rapid urine tests for phenothiazines (Forrest, Forrest, & Mason, 1961) on the day of the experiment. These tests indicated that the patients were generally taking less than the prescribed dosage. Several types of schizophrenia were represented in Group Z, 10 Ss being diagnosed paranoid, 8 undifferentiated, and 4 affective. All of the undifferentiated schizophrenics and 5

of the paranoids were considered to be chronic, while the rest were considered reactive. The time elapsed since the initial diagnosis of schizophrenia averaged 5.1 yr. The mean number of hospitalizations was 2.5, and the average length of hospitalization was 4.7 mo.

Exosomatic basal skin resistance and GSRs were monitored by a constant voltage system (Montagu & Coles, 1966) supplying 4 v. and recorded on an Offner type R dynograph. The CSs were two 8-sec. tones of 450 and 1,400 cps that were substantially above threshold. A pilot group was used to equate psychophysically the intensities of the two tones. The UCS consisted of illumination of a 6-w. jeweled white light mounted on a panel facing S at eye level from a distance of 42 in. The RT response consisted of lifting the middle finger of the preferred hand from a microswitch. The switch was located on the armrest of S's chair, and required no more than the weight of the finger for its depression. Hunter timers controlled temporal intervals, and an Erie-Pacific digital timer recorded RT in milliseconds. The RT signal was presented coincident with tone offset and was terminated by release of the switch.

Procedure

The S was seated in a semidark, sound-attenuated room adjoining the instrumentation room. The active Ag-AgCl electrode was placed on the volar surface of the second phalange of the second finger of the nondominant hand, and the reference electrode was placed on the ventral surface of the forearm. Both electrodes were coated with a gel that is isotonic with sweat. After attachment of electrodes, the S was oriented to the RT task using highly exhortative instructions stressing improvement of performance with repeated presentations. Three practice RT trials followed, tones omitted. After 2 min. of recording with S at rest, instructions for intertrial reports were presented. The S was told in effect to state her observations, thoughts, and feelings about events occurring in the experiment whenever the experimenter said "report" over the intercommunication system. No reference was made to the tones. Thereupon, 8 adaptation, 24 acquisition, and 8 extinction trials were administered, with intertrial reports interspersed between all trials. In each group, half of the Ss had the 450-cps tone paired with the RT task, and half the 1,400-cps tone. There were no test trials. Presentations of the paired tone (CS+) and unpaired tone (CS-) were regularly alternated in a single alternation pattern. The intertrial intervals averaged 51 sec. ($SD = 19$ sec.) for Group N and 54 sec. ($SD = 23$ sec.) for Group Z. Following the last extinction trial, all Ss were interviewed with a standard series of increasingly informative questions.⁴ The reports and interviews were tape-recorded and transcribed for subsequent analysis.

⁴Copies of the instructions for the RT task, instructions for intertrial reports, the postconditioning interview, and the rating schedules for the latter two have been deposited with the National Auxiliary

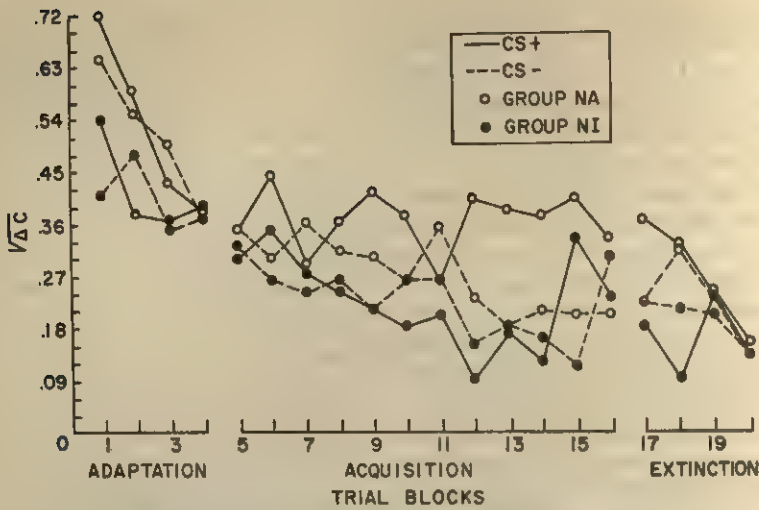


FIG. 1. Mean first-interval GSR magnitudes for accurately verbalizing normal Ss (Group NA, $N = 12$) and inaccurately verbalizing normal Ss (Group NI, $N = 10$).

RESULTS

Verbalizations

Three raters independently content analyzed the interviews for accuracy of Ss' conceptualizations of the conditional stimulus contingencies. Raters worked without knowledge of Ss' GSR conditioning. Two criteria were used to establish the accuracy of Ss' statements of stimulus relations based on either a statement that the UCS followed only one type of tone or that the UCS appeared in a regularly alternating pattern. Classification of S as an accurate verbalizer for either intertrial reports or postconditioning interviews required two ratings of accuracy not considered questionable by either rater, or three ratings of accuracy if one or more raters considered their rating questionable. For the interview ratings, verbalizations that were elicited by questions containing information about the differential stimulus contingencies or the alternating trial sequence were not used to classify Ss.

In Group N, 7 Ss were rated as having verbalized accurately during both intertrial

reports and interviews. Three Ss verbalized accurately only during the interview, and 1 only during intertrial reports. In Group Z, 1 S verbalized accurately both during reports and the interview, and 2 Ss only during the interview. Final classification of an S as an accurate verbalizer was based on either the ratings of the intertrial reports or the interviews. In Group N, 12 Ss were classified as accurate, and 10 as inaccurate. In Group Z, 3 Ss were classified as accurate, and 19 as inaccurate ($\chi^2 = 6.42$, $df = 1$, $p < .02$, two-tailed).

GSR Measures

The 8-sec. interval beginning 1 sec. after tone onset was divided into subintervals. A GSR originating 1-5 sec. after tone onset was scored as a first-interval response. A GSR originating 5-9 sec. after tone onset was scored as a second-interval response. A third-interval response was scored for a GSR originating during the interval 9-13 sec. following tone onset. Each GSR was measured from its onset to its peak and the amplitude expressed as the square root of the conductance change.

First Interval GSRs

Mean first interval GSR magnitudes are depicted in Figure 1 for the group of 12 accu-

Publications Service. Order Document No. 00526 from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Remit in advance \$1.00 for microfiche or \$3.00 for photocopies and make checks payable to: Research and Microfilm Publications, Inc.

rately verbalizing normal Ss (Group NA) and the group of 10 inaccurately verbalizing normal Ss (Group NI). The first CS+ acquisition trial was omitted from consideration because of sensitization effects of the first RT trial. Three factor analyses of variance (Groups NA versus NI, CS+ versus CS- and trial blocks comprised of pairs of successive CS+ and CS- trials) were calculated separately for the three phases of the conditioning procedure. During adaptation, Groups NA and NI did not differ significantly when responses to CS+ and CS- tones were combined ($F < 1$, $df = 1/20$) nor was the Trials \times Groups interaction significant ($F = 1.5$, $df = 3/60$). Differentiation of GSR magnitudes to CS+ versus CS- did not occur during adaptation for the combined groups ($F < 1$).

During acquisition, significantly differentiated GSR magnitudes to CS+ and CS- were obtained for Groups NA and NI combined ($F = 5.8$, $df = 1/20$, $p < .05$). The degree of conditioning differed significantly between the groups ($F = 5.5$, $df = 1/20$, $p < .05$), with conditional GSR differentiation found only in Group NA. Groups NA and NI combined did not differentiate significantly during extinction ($F < 1$), nor did the two groups differ in this respect ($F = 2.0$, $df = 1/20$). The groups did not differ in combined reactivity to CS+ and CS- during acquisition or

extinction ($F = 1.7$, $df = 1/20$; $F < 1$, respectively). It may be noted that differential GSR conditioning does not require incremental responding to CS+, but simply that there be differential responding to the CS+ and CS-.

Mean first-interval GSR magnitudes for Groups NA and NI combined (Group N) and for Group Z are shown in Figure 2. Three factor analyses of variance (Groups N versus Z, CS+ versus CS-, and trial blocks comprised of successive CS+ and CS- trials) were calculated separately for adaptation, acquisition, and extinction. During adaptation there was a nonsignificant tendency for greater reactivity in Group N ($F = 3.4$, $df = 1/42$, $p < .10$). During acquisition trial blocks the degree of differentiation between CS+ and CS- was significantly greater for Group N than for Group Z ($F = 5.8$, $df = 1/42$, $p < .05$). As shown in Figure 2, conditional differentiation of first-interval GSRs was not observed in Group Z. Because only three Ss could be classified as accurate verbalizers, assessment of GSR differentiation within Group Z in terms of the accuracy of Ss verbalizations was not attempted. Acquisition reactivity levels to CS+ and CS- combined did not differ between Groups N and Z ($F = 1.3$, $df = 1/42$). During extinction, no differentiation was observed in either group, and there were no differences in reactivity levels.

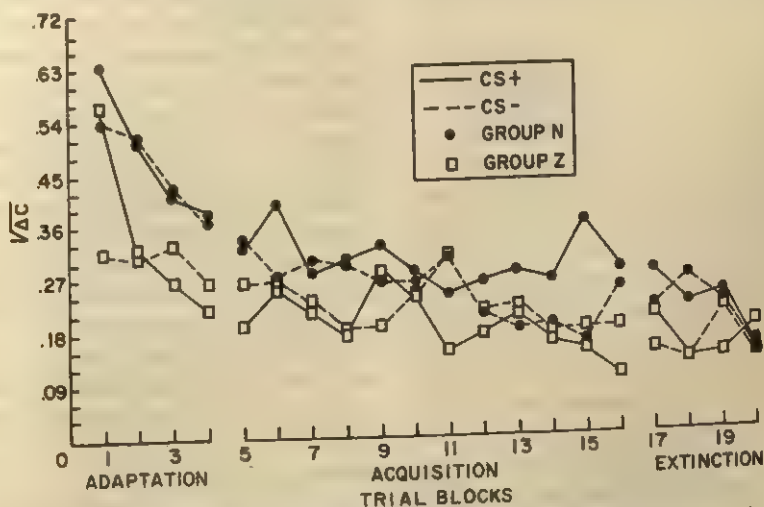


FIG. 2. Mean first-interval GSR magnitudes for normal Ss (Group N, $N = 22$) and schizophrenic Ss (Group Z, $N = 22$).

TABLE 1
MEAN SECOND-INTERVAL GSR MAGNITUDES

Group	N	Trial blocks					
		1-4 Adaptation		5-16 Acquisition		17-20 Extinction	
		CS+	CS-	CS+	CS-	CS+	CS-
NA	12	.13	.12	.12	.07	.11	.05
NI	10	.14	.07	.07	.07	.04	.08
N(NA + NI)	22	.14	.10	.09	.07	.08	.06
Z	22	.03	.03	.12	.10	.07	.09

Second- and Third-Interval GSRs

Second-interval magnitudes are presented in Table 1. The distribution of these data suggested the use of nonparametric analyses. Two-tailed tests of significance were used throughout. For each S, second-interval amplitudes were summed over trials separately for CS+ and CS- within each phase of the experiment. During adaptation, Groups NA and NI did not differ when responses to the CS+ and CS- tones were combined ($U = 55$, $df = 12/10$), nor did either Groups NA or NI demonstrate GSR differentiation ($T = 13$, $df = 9$; $T = 3$, $df = 5$, respectively). During acquisition, Groups NA and NI combined showed a statistically nonsignificant tendency toward differentiation ($T = 67.5$, $df = 21$, $p < .10$). When Groups NA and NI were considered separately, GSR differentiation ap-

proached statistical significance for Group NA ($T = 15$, $df = 12$, $p = .06$) but not for Group NI ($T = 21$, $df = 9$). When Groups NA and NI were compared directly, it was found that the two groups did not differ significantly in differentiation ($U = 37$, $df = 12/10$) or in reactivity to CS+ and CS- combined ($U = 56$, $df = 12/10$). During extinction, Groups NA and NI combined did not differentiate ($T = 54$, $df = 17$). Significant differentiation was found in Group NA ($T = 10$, $df = 11$, $p < .05$) but not in Group NI ($T = 5$, $df = 6$). Direct comparison of the groups revealed greater differentiation for Group NA ($U = 23$, $df = 12/10$, $p < .02$).

Analysis of second-interval GSRs for Group Z produced no statistically reliable results. The schizophrenic Ss did not differentiate during acquisition ($T = 64$, $df = 20$), nor did comparison of the differentiation of Groups N and Z differ significantly ($U = 236$, $df = 22/22$). Furthermore, Groups N and Z did not differ in reactivity to CS+ and CS- combined ($U = 214$, $df = 22/22$). During extinction, Group Z did not differentiate ($T = 82$, $df = 16$), and Groups N and Z did not differ in either differentiation ($U = 175$, $df = 22/22$) or reactivity ($U = 236$, $df = 22/22$).

Mean third-interval GSR magnitudes during extinction trials were Group NA, CS+ = .12, CS- = .12; Group NI, CS+ = .06, CS- = .10; Group Z, CS+ = .12, CS- = .10. There was no tendency for these results to be statistically significant for any of the groups.

Reaction Time

The RT curves for Groups NA, NI, and Z are displayed in Figure 3. In each group, RT

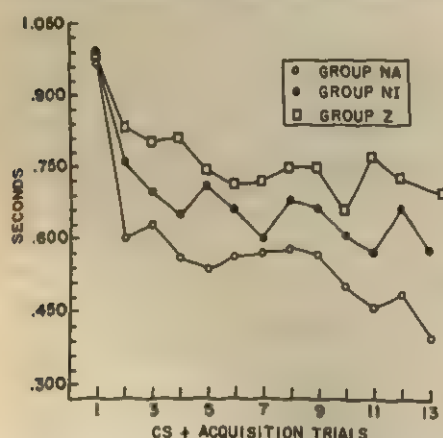


FIG. 3. Mean RT for the accurately verbalizing normal Ss (Group NA, $N = 12$), the inaccurately verbalizing normal Ss (Group NI, $N = 10$), and the schizophrenic Ss (Group Z, $N = 22$).

decreased from early to late trials. The groups did not differ in RT on the first trial, nor in extent of reduction of RT from early to late trials. Mean RT over the final four acquisition trials differed significantly between Groups N and Z ($U = 98$, $df = 22/22$, $p < .02$) and between Groups NA and NI ($U = 12$, $df = 12/10$, $p < .002$). A similar comparison between Groups NI and Z was not significant ($U = 69$, $df = 10/22$).

To determine a possible relationship between conditional GSR differentiation and RT in Group N, rank-order correlations were performed. A GSR differentiation score was computed by subtracting the summed GSR amplitudes during the last four CS- trials from the summed GSR amplitudes during the last four CS+ trials. Between the first-interval GSR differentiation score and mean RT during the last four trials ($r = -.38$, $df = 22$, $p < .05$) and between the second-interval GSR differentiation score and mean RT ($r = -.48$, $df = 22$, $p < .05$). Failure to obtain conditioning in Group Z precluded similar correlational analyses in that group.

Several supplementary correlations were computed to ascertain whether RT was related to GSR indexes other than the differentiation score. Considering the final four trials, mean RT was not significantly correlated with the first-interval GSR amplitudes to CS+ ($r = .29$, $df = 22$) nor with second-interval amplitudes to CS+ ($r = .13$, $df = 22$). The magnitude of third-interval GSRs accompanying RT was not significantly correlated with mean RT during the final four trials ($r = -.22$, $df = 22$).

DISCUSSION

The present results indicate that GSRs of normal Ss can be differentially conditioned using an RT task as the UCS. During the acquisition phase, statistically significant differentiation was obtained for first-interval GSRs, and a tendency toward differentiation was observed for second-interval GSRs. During extinction, differentiation of both first- and second-interval responses was abolished, and there was no evidence for third-interval differentiation. The present procedure may be particularly attractive in studying GSR conditioning of Ss such as children and the elderly

for whom the use of a noxious UCS is undesirable.

Certain characteristics of the present procedure may have facilitated conditioning. Instructions for the RT task were highly exhortative, presumably enhancing the Ss' attention to experimental events and their motivation to perform the RT task. Since the two conditional stimuli were presented in a single alternation pattern, conditional differentiation could be based on the alternation of the tones as well as on the differences in frequency between them. The insertion of inter-trial reports may have facilitated conditioning and increased the number of accurate verbalizers. Such effects had been previously observed using shock as the UCS (Fuhrer & Baer, 1966).

Both direct and indirect comparisons of GSR differentiation between normal Ss who could accurately verbalize conditional stimulus relations (Group NA) and those who could not (Group NI) showed reliable conditioning in Group NA but not in Group NI. During the acquisition phase there was both first- and second-interval GSR differentiation for Group NA. During the extinction phase, second-interval differentiation by Group NA was not abolished, although first-interval differentiation was. These findings are generally in accord with those obtained when a shock UCS was used (Baer & Fuhrer, 1968; Fuhrer & Baer, 1965, 1966). The concordance between conditional GSR differentiation and accurate verbalizations of stimulus relations evidently supersedes effects attributable to the type of UCS, aversive or nonaversive.

The schizophrenic group did not show differential GSR conditioning in any of the response intervals during either acquisition or extinction trials. Our results using an RT task as the UCS appear consistent with those obtained previously using an aversive UCS. Peters and Murphree (1954) report some evidence that schizophrenics show less differential conditioning of the GSR than normals using extended interstimulus intervals and either shock or a claxon as the UCS. The more recent studies of the simple GSR conditioning of schizophrenics and normals have not yielded consistent findings. Pishkin and Hershisier (1963) failed to find simple GSR

conditioning in a schizophrenic group, while Howe (1958) found that the GSR conditioning of schizophrenics and normals did not differ.

The conditionability of schizophrenics may also reflect the clinical population being sampled, though this was not evident in the present sample. Individual differences in first- and second-interval GSR differentiation were scrutinized within the schizophrenic group, but there seemed to be no systematic differences between Ss classified as chronic versus reactive or paranoid versus nonparanoid.

As predicted, it was found that the frequency of inaccurate verbalizers was associated with the presence of psychopathology. Of the 22 schizophrenics, 19 could not describe the conditional stimulus relations during either intertrial reports or the interview. In the normal group of 22 Ss, only 10 failed to verbalize accurately. The failure of the schizophrenics to verbalize accurately did not appear attributable to a marked attentional deficit or to their inability to communicate coherently. The patients were able to communicate well enough to be able to maintain themselves in the community under outpatient care. Pathognomonic types of bizarre or autistic verbalizations were rare. Finally, content analysis of verbal reports elicited between adaptation trials indicated no difference between patients and normals in the number of references made to the tones and the impending RT task. This finding suggests that the patients were capable of verbalizing attention to experimental events.

The generally convergent results provided by the schizophrenic and normal groups have implications for understanding the differential conditioning of the human adult as well as for interpreting the deficient conditionability of schizophrenics. Consistent with the results of our previous studies, an absence of conditional GSR differentiation was associated with inability to recognize the differential stimulus relations. Differential GSR conditioning was only manifested by accurately verbalizing normal Ss. It was not possible to assess the conditioning of accurately verbalizing schizophrenics since there was only three of these Ss and their GSR differentiation scores varied markedly. A characterization of the classical

conditioning process as a relatively primitive, automatic form of learning seems inadequate to account for this pattern of results. The differential classical conditioning of autonomic responses in the human adult appears to be closely related to a conceptual process reflected by the ability to verbalize stimulus relations accurately. When such a conceptualization fails to occur, conditional differentiation is not observed. Of course, it cannot be asserted that this conceptual process fails to occur for the same reasons for inaccurately verbalizing schizophrenics and normals. It does seem consistent with this viewpoint that the conditional GSR differentiation of the schizophrenic group was found to be markedly deficient compared to the normal group. Impairments of conceptual activity are a well-documented characteristic of schizophrenics and seem to have been reflected in a noticeably higher incidence of inaccurate verbalizers within this group.

The heuristic value of considering parallels between classical conditioning and RT procedures has been pointed out previously (e.g., Behar & Adams, 1966; Woodworth, 1958). The present study provided the unusual opportunity to combine these procedures, permitting a joint comparison among indexes of conditional GSR differentiation, ability to recognize stimulus relations, and RT. A particularly salient finding in the normal group was that accuracy of verbalization and GSR differentiation were both mirrored in RT. That is, the RT of accurate verbalizers was significantly faster than inaccurate verbalizers, and differential GSR conditioning scores correlated significantly with the RT. Comparison of the normal and schizophrenic groups yielded a similar pattern of findings. The RT of the schizophrenic group was significantly longer than that of the normal group, and the schizophrenics were characterized by inaccurate verbalization of stimulus relations and an absence of conditional GSR differentiation.

This pattern of results suggests that recognition of the CS-UCS relations by the accurately verbalizing normal Ss permitted these individuals to use the CS+ tone as a ready signal for making the RT response. In accord with previous studies reviewed by Woodworth and Schlosberg (1954), availability of a ready

signal was reflected in shorter RTs than shown by either inaccurately verbalizing normals or schizophrenics. Perception of the CS+ tone as a ready signal also allowed the accurately verbalizing normal Ss to adopt a preparatory set for making the RT response. Such a set probably embodies a complex pattern of physiological changes including increased muscle tension (e.g., Davis, 1940) as well as the differentiated GSR activity revealed in the present study.

It has been suggested that fundamentally different processes underlie the conditioning of first-, second-, and third-interval GSRs. For example, Gale and Stern (1967) interpret the differentiation of first-interval GSRs as conditioning of the orienting response to onset of the CSs. Second-interval GSRs are viewed as anticipatory responses representing preparation for delivery of the UCS. If responses in these intervals reflect different underlying mechanisms, GSR activity within each should bear a different functional relationship with other variables. In the present study, however, both first- and second-interval GSR differentiation for the normal group was similarly related to the accuracy of Ss' verbalizations. Furthermore, the schizophrenic group showed no GSR differentiation in any of the intervals, and both the first- and second-interval GSR differentiation for the normal group was similarly correlated with RT. These results do not justify a sharp distinction between the processes underlying GSR activity in the various intervals.

There is a remaining question concerning possible effects of the phenothiazine medication upon results obtained for the schizophrenic group. It is difficult to assess the possibility that these agents affected the relevant cognitive functioning of Ss in this group. Chapman and Knowles (1964) have shown that phenothiazine treatment has the dual effects of improving conceptual activity and decreasing alertness in these patients. Other studies indicate that phenothiazine treatment improves intelligence test performance in catatonic patients (Gilgash, 1961), but not in newly admitted schizophrenics (Judson & MacCasland, 1960). It is unlikely that drug effects were responsible for the depressed RTs of the schizophrenics. Neither Heilizer

(1959) nor Sibilio, Andrew, Dart, Moore, and Stehman (1959) found significant alterations of schizophrenics' RTs due to phenothiazine treatment. Finally, there is no available evidence that these drugs affect the conditional GSR differentiation of schizophrenics. While there is an indication that phenothiazines affect GSR indexes such as the frequency of nonspecific GSRs in schizophrenics (Bernstein, 1967), Crider and Tursky (1967) have shown in normal Ss that the frequency of nonspecific skin potential responses bears no relationship to their conditional differentiation.

REFERENCES

- BAER, P. E., & FUHRER, M. J. Cognitive processes during differential trace and delayed conditioning of the GSR. *Journal of Experimental Psychology*, 1968, 78, 81-88.
- BEHAR, I., & ADAMS, C. K. Some properties of the reaction-time ready-signal. *American Journal of Psychology*, 1966, 79, 419-426.
- BERNSTEIN, A. S. Electrodermal base level, tonic arousal, and adaptation in chronic schizophrenics. *Journal of Abnormal Psychology*, 1967, 72, 221-232.
- CHAPMAN, J., & MCGHIE, A. A comparative study of disordered attention in schizophrenia. *Journal of Mental Science*, 1962, 108, 487-500.
- CHAPMAN, L. J., & KNOWLES, R. R. The effects of phenothiazine on disordered thought in schizophrenia. *Journal of Consulting Psychology*, 1964, 28, 165-168.
- CRIDER, A., & TURSKY, B. Relationship of electrodermal lability and reflex sensitivity to differential conditioning. *Psychonomic Science*, 1967, 9, 225-226.
- DAVIS, R. C. Set and muscular tension. *Indiana University Publications in Science Series*, 1940, No. 10.
- FORREST, F. M., FORREST, I. S., & MASON, A. S. Review of rapid urine tests for phenothiazine and related drugs. *American Journal of Psychiatry*, 1961, 118, 300-307.
- FUHRER, M. J., & BAER, P. E. Differential classical conditioning: Verbalization of stimulus contingencies. *Science*, 1965, 150, 1479-1481.
- FUHRER, M. J., & BAER, P. E. Effects of intertrial reports on cognitive and GSR differentiation of conditional stimuli. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1966, 1, 53-54.
- GALE, E. N., & STERN, J. A. Conditioning of the electrodermal orienting response. *Psychophysiology*, 1967, 3, 291-301.
- GILGASH, C. A. Thorazine therapy with catatonic schizophrenics in relation to Wechsler verbal and performance subtest comparison. *Journal of Clinical Psychology*, 1961, 17, 95.
- HEILIZER, F. The effects of chlorpromazine upon psychomotor and psychiatric behavior of chronic

- schizophrenic patients. *Journal of Nervous and Mental Disease*, 1959, 128, 358-365.
- HOWE, E. S. GSR conditioning in anxiety states, normal and chronic functional schizophrenic subjects. *Journal of Abnormal and Social Psychology*, 1958, 56, 183-189.
- JUDSON, A. J., & MACCASLAND, B. W. The effects of chlorpromazine on psychological test scores. *Journal of Consulting Psychology*, 1960, 24, 192.
- LANG, P. J., & BUSS, A. H. Psychological deficit in schizophrenia: II. Interference and activation. *Journal of Abnormal Psychology*, 1965, 70, 77-106.
- MARQUIS, D. G., & PORTER, J. M. Differential characteristics of conditioned eyelid responses established by reflex and voluntary reinforcement. *Journal of Experimental Psychology*, 1939, 24, 347-365.
- MARTIN, I. Discriminatory GSRs. *Activitas Nervosa Superior*, 1965, 7, 217-223.
- MONTAGU, J. D., & COLES, E. M. Mechanism and measurement of the galvanic skin response. *Psychological Bulletin*, 1966, 65, 261-279.
- PETERS, H. N., & MURPHREE, O. D. The conditioned reflex in the chronic schizophrenic. *Journal of Clinical Psychology*, 1954, 10, 126-130.
- PISHEKIN, V., & HERSHISER, D. Respiration and GSR as functions of white sound in schizophrenia. *Journal of Consulting Psychology*, 1963, 27, 330-337.
- SIBILIO, J. P., ANDREW, O., DART, D., MOORE, K. B., & STEHMAN, V. A. Effects of promazine hydrochloride on attention in chronic schizophrenia. *AMA Archives of Neurology and Psychiatry*, 1959, 81, 114-120.
- WOODWORTH, R. S. *Dynamics of behavior*. New York: Holt, 1958.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.
- ZAHN, T. P., ROSENTHAL, D., & SHAKOW, D. Effects of irregular preparatory intervals on reaction time in schizophrenia. *Journal of Abnormal and Social Psychology*, 1963, 67, 44-52.

(Received July 15, 1968)

NEW IN 1969

DEVELOPMENTAL PSYCHOLOGY

Editor: Boyd R. McCandless

Developmental Psychology contains articles which represent the broad range of growth and development and their major associated variables. Chronological age as well as sex, socioeconomic status and efforts of physical growth variables (including time of maturation and body build) are all considered relevant development variables.

Although the adolescent and the aging population appear to deserve particular attention at this time, *Developmental Psychology* indeed covers the entire span. Cross-species articles and articles concerning developmental research with retardates are also included.

The articles in *Developmental Psychology* are not limited to any one approach such as "experiments versus studies" or "cognition versus motivation".

**Bimonthly. One volume per year.
Subscription Price: \$10.00 per year**

**AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036**

Journals of the AMERICAN PSYCHOLOGICAL ASSOCIATION

Subscription Rates for Calendar Year 1969

AMERICAN PSYCHOLOGIST. Official papers of the Association and articles on Psychology. Monthly. The 1969 volume is Vol. 24. Subscription: \$10.00 (foreign \$10.50). Single copy \$1.00.

CONTEMPORARY PSYCHOLOGY. Critical reviews of books, films, and research material in the field of psychology. Monthly. The 1969 volume is Vol. 14. Subscription: \$10.00 (foreign \$10.50). Single copy \$1.00.

JOURNAL OF ABNORMAL PSYCHOLOGY. Theoretical and research conditions in behavior pathology. Bimonthly; first issue appears in February. The 1969 volume is Vol. 74. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00. (Continues volume sequence of the former Journal of Abnormal and Social Psychology.)

JOURNAL OF APPLIED PSYCHOLOGY. Applications of psychology in business and industry. Bimonthly; first issue appears in February. The 1969 volume is Vol. 53. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00.

JOURNAL OF COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY. Original contribution in the field of comparative and physiological psychology. Beginning in 1969: Monthly. Three volumes per year, four issues in each volume. The 1969 volumes are Vols. 67, 68 and 69. Subscription: \$30.00 for 3 vols. (foreign \$30.50). Single copy \$3.00.

JOURNAL OF COUNSELING PSYCHOLOGY. A primary publication medium for research in counseling theory and practice. Bimonthly; first issue appears in January. The 1969 volume is Vol. 16. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00.

JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY. Research in clinical psychology: psychological diagnoses, psychotherapy, personality, psychopathology. Bimonthly, first issue appears in February. The 1969 volume is Vol. 33. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00.

JOURNAL OF EDUCATIONAL PSYCHOLOGY. Studies of learning and teaching; measurement of psychological development, psychology of school subjects, methods of instruction, school adjustment. Bimonthly; first issue appears in February. The 1969 volume is Vol. 60. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY. Original contributions of an experimental character. Monthly. Beginning 1969: Four volumes per year, three issues in each volume. The 1969 volumes are Vols. 79, 80, 81 and 82. Subscription: \$40.00 for 4 vols. (foreign \$40.50). Single copy \$4.00.

JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY. Theoretical and research papers on personality dynamics, group process, and the psychological aspects of social structure. Monthly. Three volumes per year, four issues in each volume. The 1969 volumes are Vols. 11, 12 and 13. Subscription: \$30.00 for 3 vols. (foreign \$30.50). Single copy \$3.00.

PSYCHOLOGICAL ABSTRACTS. Noncritical abstracts of the world's literature in psychology and related subjects. Monthly. Approximately 20,000 abstracts published annually. The 1969 volume is Vol. 43. Subscription: \$40.00 (foreign \$40.50). Single copy \$4.00.

PSYCHOLOGICAL BULLETIN. Evaluative reviews of research literature, discussions of research methodology in psychology. Monthly. Two volumes per year, six issues in each volume. The 1969 volumes are Vols. 71 and 72. Subscription: \$20.00 for 2 vols. (foreign \$20.50). Single copy \$2.00.

PSYCHOLOGICAL REVIEW. Original contributions of a theoretical nature. Bimonthly; first issue appears in January. The 1969 is Vol. 76. Subscription: \$10.00 (foreign \$10.50). Single copy \$2.00.

NEW IN 1969 DEVELOPMENTAL PSYCHOLOGY. Studies on growth and development and their associated variables. Bimonthly; first issue appears in January. The 1969 volume is Vol. 1. Subscription \$10.00 (foreign \$10.50). Single copy \$2.00.

Discounts: AGENTS and BOOK DEALERS receive a 10% discount on all orders. There is no club arrangement. There are no quantity, cash, library, or other special discounts. *Payment in U.S. funds.*

★ Subscriptions are available on a January through December basis only ★

AMERICAN PSYCHOLOGICAL ASSOCIATION
Subscription Department
1200 Seventeenth Street, N.W.
Washington D. C. 20036

Journal of Abnormal Psychology

Diary No. 269
Date 12-6-71
File No. lib
Ednl. Psy. Assoc.

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editors: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

ARTHUR C. CARR, *New York State Psychiatric Institute and Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

RICHARD L. BLUMENTHAL
*New York State Psychiatric Institute
and Columbia University*

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JOSEPH L. FLEISS
*New York State Department of Mental
Hygiene and Columbia University*

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MARGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

* The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt

Journal of Abnormal Psychology

Department of Psychology

N. Y. State Psychiatric Institute & Columbia University

722 West 168th St.

New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR	MYRTLE HENDRICKS	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

FREQUENCY OF PSYCHOTHERAPEUTIC SESSION AS A FACTOR AFFECTING OUTCOME: ANALYSIS OF CLINICAL RATINGS AND TEST RESULTS¹

CHRISTOPH M. HEINICKE²

Reiss-Davis Child Study Center, Los Angeles, California

Two groups of boys referred for learning disturbances were seen either in once-a-week or in four-times-a-week psychoanalytic psychotherapy. Children seen more frequently showed greater ego flexibility by the end of treatment, as well as a more adequate level of ego integration during the follow-up period. Analysis of test results indicated that these same children improved at a faster rate in tests of reading, spelling, and vocabulary, and responded more flexibly to the Rorschach cards.

The importance of the frequency with which a patient is seen in psychotherapy has long been recognized, but few systematic studies are available. Imber, Frank, Nash, Stone, and Gliedman (1957) found that adult patients evaluated after 6 mo. showed greater gains in social effectiveness if seen once a week as opposed to every other week. By contrast, Lorr, McNair, Michaux, and Raskin (1962) found that the variation of every other week, versus once a week, versus twice a week, made little difference in the

status of the patient 4 and 8 mo. after treatment was initiated. For patients in treatment at least a year, however, those seen on a more frequent basis "describe themselves as more outspoken, assertive, independent, and determined to protect their own interests." A follow-up of a part of this initial sample (McNair, Lorr, Young, Roth, & Boyd, 1964) found significant relations between the number of interviews and 4 of the 13 outcome criteria.

The present study compared two groups of boys in terms of treatment and outcome variables. One group was seen once a week and the other group four times a week. Two psychoanalytic child therapists with similar training and experience treated the children. All the mothers were seen on a once-a-week basis, and, where appropriate, the fathers also had contact with a therapist. Using certain broad diagnostic considerations, cases were assigned to matched samples on a case-to-case basis. All treatments were terminated at the request of the parents and with the consent of the therapist. Information on the process variables was derived from the record dictated by either the child's or parent's therapist immediately after the sessions, but is not discussed in this paper. Information on the outcome was derived from multiple assessments at the beginning, end, and 1 and 2 yr. after the end of treatment.

The total data available on each child for each assessment point were initially organized in terms of the Diagnostic Profile (Freud, 1962). Further analysis of the outcome data

¹ This study was planned while the author was a Fellow at the Center for the Advanced Study in the Behavioral Sciences, and was initiated at Mount Zion Hospital, San Francisco, under grants from the Rosenberg Foundation and the National Institutes of Health, Grant M-2948. In addition to the author, the following were members of the project: Joseph Afterman, Marian Bradley, Leah Kaplan, Anneliese F. Korner, and Jean Moore. The study is being completed at Reiss-Davis Child Study Center, Los Angeles, with the assistance of a grant from the Grant Foundation. The author particularly wishes to express his gratitude to Walter Raine, who advised and assisted on all statistical analyses, and to Barbara Carr, and Heiman Van Dam, who aided in testing the reliability of the clinical ratings. Thanks to Rocco L. Motto, Norman Reider, Anna Maenchen, Mortimer Meyer, Elise Greenhouse, and Dorothy Habben, for their generous help in carrying out this project. Without the support of the school systems involved, this study would not have been possible. Finally, the author is much indebted to Sheila Spellman, Rosemary Ginn, Diana Hager, Roselyn Katz, and Stella Gunther for their devoted assistance.

² Requests for reprints should be sent to Christoph M. Heinicke, Reiss-Davis Child Study Center, 9760 West Pico Boulevard, Los Angeles, California 90035.

involved clinical ratings of the Profile material as well as test scores. It is the purpose of this paper to show that the results of this analysis are consistent with previously formulated clinical hypotheses (Heinicke, 1965a). The most general of these hypotheses was that the children seen four times as opposed to once a week make greater progress during the follow-up period in the level of ego integration, flexibility, and adaptation. It was anticipated that this differential growth would be reflected not only in clinical ratings of these attributes but also in IQ scores, Rorschach responses, and the rate of improvement in academic subjects.

METHOD

Methods of Assessing the Children

At the beginning and end of treatment, the following data were available: interviews conducted by therapists with the parents and teachers, the sessions of therapist and child, one or two interviews with a child psychiatrist, and the results of extensive testing with the child (Revised Stanford-Binet, Form L, the Wide Range Achievement Test, the Rorschach, parts of the TAT and Michigan Picture Test, and the Draw-a-Person Test).

This same information was again available at the first and second follow-ups, but the child was not seen again by the diagnostic psychiatrist, and the second follow-up included no testing. Scores on reading tests were available from the school records.

Subjects

Data analysis, including group comparison, was carried out on two samples of four boys each who (a) ranged in age 6 yr. 8 mo.-10 yr. 5 mo.; (b) were referred primarily for a learning disturbance associated with psychological difficulties; (c) were threatened with being or had already been held back in school; (d) were unable to profit from supplemental educational help; (e) achieved a Stanford-Binet IQ of 90 or better; (f) came from intact business or professional families; and (g) were in treatment from 1½ to 2½ yr. Case-to-case matching insured that the two groups did not differ in regard to the above characteristics. They also did not differ at the beginning of treatment on any of the ratings which differentiated them after treatment.

Ratings Based on the Diagnostic Profile

The Diagnostic Profile provides a common framework and language in narrative form, for making inferences from multiple assessments of the child and his family. A full example is given in Heinicke (1965a).

The forty-five 10-point ratings used to describe the children reflect a number of important theoretical

distinctions suggested by the Profile Outline, as well as the nature of the problem being studied. The examples of the ratings cited below are central to the results which emerged from the data analysis. In addition to much more specific definitions, characteristics of Points 10, 5, and 1 for each rating were specified descriptively.³

1. Child's level of ego integration—To what extent is the child's functioning characterized by unity? To what extent do important areas of functioning not contribute to his adaptation and how effective is that adaptation?

2. Extent to which the child's defenses are balanced—To what extent is the defensive organization characterized by the use of many of the important mechanisms or the excessive use of single ones? (Freud, 1962).

3. Child's capacity to express a variety of affects—To what extent are a variety and balanced use of affects available to the child?

4. Capacity to form peer relationships—What is the number and quality of the peer relationships which the child can initiate and sustain?

5. Extent of development from dependency to emotional self-reliance—Where along the line from the newborn's utter dependence on maternal care to the young adult's self-reliance should the child be placed? (Freud, 1963).

Reliability of the ratings. The reliability of the ratings by the therapist was checked in three ways. The first was to correlate the therapist's ratings with those done on the same Profiles by another clinician. The median Q-type coefficient was .85; the *N* for each computation was 45 or the number of ratings. A second check was to correlate the ratings based on the raw clinical data. The median coefficient between the therapist and psychiatrist was .75. A third check was to estimate the reliability of the factor score.

Analysis of Test Scores

All test results were initially analyzed clinically (Heinicke, 1965a). Individual Rorschach score summaries were later derived by a person not familiar with the frequency of treatment each child experienced.

The grade level scores in reading, spelling, and arithmetic derived from the Wide Range Achievement Test were correlated with the grade placements made by the teachers. The median correlation was .70 ($p < .05$).⁴

³ The complete list of the definitions of the 45 ratings and details of the subsequent analysis may be obtained by writing to the author.

⁴ Rate of change was used in all further data analysis. For the beginning of treatment the gain made up to that point was divided by the number of months the child had been in school. Given the low association between the rate of improvement before treatment and that during the first year after

RESULTS

Group Differences on Factor Scores

There were no significant group differences in group mean ratings at the beginning of treatment, but three statistically significant differences indicating greater ego flexibility in the children seen four times a week emerged by the end of treatment; 39 of the 45 dimensions showed a statistically significant difference by the first or second follow-up points. Since many ratings were intercorrelated, a principal components factor analysis (with 1s in the diagonal and with rotation of all factors with an eigenvalue greater than 1) was carried out on the table of intercorrelation of the 45 ratings computed for each of the four assessment points.

The Four Factors

Of the factors isolated, four were obtained for each assessment: (a) level of ego integration, (b) extent of ego flexibility, (c) capacity for forming peer relationships, and (d) extent of self-reliance. Those ratings associated with a factor to the extent of .625 or better at three of the four assessments points further defined the factor.

From the first factor, level of ego integration, which at the various assessment points accounted for 50-69% of the variance, the following rating items remained after the nonsignificant contributions had been eliminated: level of ego integration, capacity to sublimate, absence of primary process, effectiveness of the defenses, capacity to tolerate anxiety, level of self-esteem as a function of the child's achievement, ability to assert himself, extent of academic ambition, ability to persist in a task, realistic concern about academic progress, absence of denial or avoidance of the academic problem, progress made in libidinal development, predominance of progressive over regressive forces, and the congruence between various developmental lines.

The factor of ego flexibility accounted for 2-17% of the variance. Four ratings loaded on the factor to the extent of .625 or better

treatment, as well as the high reliability of the test scores, it was not felt necessary to compute true gain scores (Lord, 1956).

TABLE 1

SUM OF FACTOR SCORES FOR CHILDREN IN ONCE AND FOUR-TIMES-A-WEEK THERAPY

Factor	Point of treatment			
	Beginning	End	First follow-up	Second follow-up
Level of ego integration				
Once a week	10.1	4.7	-16.0	-24.8
Four times a week	-10.1	-4.7	16.0	24.8
Extent of ego flexibility				
Once a week	4.7	-2.3	-5.3	-7.3
Four times a week	-4.7	2.3	5.3	7.3
Capacity for peer relations				
Once a week	1.8	.8	-1.3	-6.3
Four times a week	-1.8	-.8	1.3	6.3
Extent of self-reliance				
Once a week	1.1	-.4	-1.3	-1.1
Four times a week	-1.1	.4	1.3	1.1

* Difference significant at $p < .05$.** Difference significant at $p < .01$.

on at least three of the assessment occasions: capacity to express a variety of affects, extent of balance in the defense organization, negatively with the extent of repression of aggression, and nondefensive use of humor.

The factor of the capacity to form peer relationships accounted for 5-10% of the variance and included: capacity to form peer relationships, development from egocentricity to companionship, and level of self-esteem as related to the body image.

The fourth factor, self-reliance, was again named on the basis of several ratings and accounted for 4-7% of the variance at each of the assessment points. However, only one rating, development from wetting and soiling to autonomous bladder and bowel control, was loaded on this factor at three of the assessment points.

Group differences on four factors. The above procedure made it possible to derive a score for each child for each factor at each assessment point. Differences in group trends could thus be analyzed. Using the ratings of the total sample of eight children, standard

TABLE 2

p VALUES DERIVED FROM ANALYSIS OF VARIANCE OF FOUR FACTORS

Source	Ego integration	Ego flexibility	Peer relationships	Self-reliance
Frequency	<i>ns</i>	.05	<i>ns</i>	<i>ns</i>
Frequency \times Assessment Point	.005	.005	.01	<i>ns</i>

scores ($M = 0$, $SD = 1$) were computed separately for each rating at each assessment point. This procedure allowed summing the scores of a given child across the various ratings within a factor. Thus a "factor score"⁵ was derived. This procedure also results in the group mean for all eight children being zero, and since the total group was split into two (once a week vs. four times a week), the two subgroups necessarily had scores that together summed to zero. Table 1 gives the sum of the factor scores for each of the two groups for each assessment point.

To assess the statistical significance of the difference in the trends for the children seen once versus four times a week, an analysis of variance (trend analysis, Edwards, 1950) was carried out on each factor. The *p* values resulting from this analysis are summarized in Table 2. Most important, it shows that there is a statistically significant difference between the two frequency-of-treatment groups on three of the factors when the assessment point is taken into account (Frequency \times Assessment Point). This in turn warranted the application of *t* tests to these three factors. In Table 1 it can be seen that the children seen four times a week showed, by the second follow-up, a more adequate level of ego integration, ego flexibility, and capacity to form peer relationships. Furthermore, this same group was already more flexible in its ego functioning by the time of the first follow-up.

The possible effect of the previous level of the rating on subsequent assessment was taken into account by computing the follow-up difference scores: beginning to end of

⁵ This factor score is an approximation since no differential weightings were used. The composite score is similar to a test whose items each have unit weight.

treatment, end to first follow-up, and first to second follow-up. Using a *t* for difference scores, the two frequency groups could then be contrasted for each factor. The children seen four times a week, when compared with the once a week children, showed significant gains in (a) extent of ego flexibility, from the beginning to the end of treatment; (b) level of ego integration, from the end to the first follow-up; and (c) capacity for peer relations, from the first to the second follow-up. The *p* values, respectively, were .02, .001, and .02.⁶ This analysis, then, again indicates significant group differences for three of the factors, and also pinpoints the time interval when the children seen more frequently made particularly significant gains.

Reliability of "factor scores." The nature of the data analysis allowed application of Lindquist's formula (Lindquist, 1953, Formula No. 152, p. 361) to estimate the reliability of the "factor" or composite score (see Table 3). Reliability here means the intraitem agreement after partialing out individual differences.

Ego Integration, Total IQ, IQ Change, and Environmental Impact

The association of the factor scores with variables independent of the clusters was also studied. The ego integration factor scores correlated with the Stanford-Binet IQ scores at the beginning, end, and first follow-up point to the extent of .96, .91, and .66 (all *p*'s $< .01$). Moreover, the increase in IQ from the beginning to the first follow-up correlated to the extent of .83 ($p < .01$) with the level of ego integration attained at the first follow-up point.⁷

The ego integration scores at the beginning of treatment also correlated significantly ($p < .05$) with the rate of achievement in reading, arithmetic, and spelling, and negatively with ratings of extent to which the

⁶ Since the regression of the factor scores was not the same for the two groups, analysis of covariance was not appropriate. It was nevertheless carried out and the results are consistent with those reported above.

⁷ The absolute level of IQ at beginning of treatment did not correlate significantly with the increment in IQ from beginning to follow-up point.

child's environmental influences impede his potential for development.² Significant associations were not found for other assessment points.

Group Differences in Rate of Academic Achievement

The rate of improvement in the various subjects for the children seen once and four times a week is given in Table 4; rate of improvement for reading has been graphed in Figure 1. By contrast to the other two subjects, there were no statistically significant group differences for the subject of arithmetic (the Mann-Whitney *U* test was used for all the analyses of test scores).

Particularly striking in these results is the reversal of group trends in reading: the once-a-week children showed a greater improvement during the first year of treatment, but then their growth decelerated, especially right after treatment.

Group Differences in Total IQ and Vocabulary Subtest

The median total IQ scores for the two groups are given in Table 5. The children seen once a week by comparison with those seen more frequently showed a significant drop at the end of treatment ($p \cong .029$) but in the year after returned to the level seen at the beginning.

While the median IQ scores were thus not different for the two groups *after* treatment, their performance on the vocabulary subtest did differ. Comparison of the increments from the beginning to the end of treatment and from the end to the first follow-up favored the children seen four times a week (the respective p values were .029 and .057).

Group Differences on Rorschach Indexes of Ego Flexibility

As part of the effort to develop measures of outcome independent of the clinical judg-

² To derive an indication of environmental impact on the child is clearly difficult. The number of quantitative and qualitative distinctions is vast. The raters were able to differentiate the extent to which the environmental influences impede the child's potential for development. Only those features of the environment were considered which do in fact make an impact.

TABLE 3
RELIABILITY OF "FACTOR SCORES" AT EACH POINT
IN TREATMENT

Factor	Point of treatment			
	Begin- ning	End	First follow- up	Second follow- up
Ego integration	.903	.931	.944	.924
Ego flexibility	.852	.764	.845	.802
Peer relationships	.873	.708	.828	.767

ment of either tester or therapist, the two groups were also compared in terms of quantitative Rorschach indexes reflecting differences in (a) the degree of constrictive control and (b) emotional responsiveness.

The following were considered in relation to the degree of constrictive control: the number of determinants used, the children's percentage of form responses to the Rorschach ($F\%$), the percentage of good form responses ($F+$), and the percentage of animal responses ($A+Ad\%$; Klopfer, Ainsworth, & Klopfer, 1954). It was found that the children seen four times a week used a greater number of different determinants at first follow-up; a median number of 6 as opposed to 4 ($p \cong .03$). Their lower median $F\%$ (31 vs. 46%) is consistent with this finding, and since their median $F+\%$ was 90% and thus close to the 93% expected for their age, the below average $F\%$ and greater use of a variety of determinants is more likely an indicant of spontaneity rather than a break in reality appreciation. The latter conclusion is supported by the fact that these children showed an increase from the end of treatment to the first follow-up point in the number of popular responses ($p \cong .01$).

There were other indications that the once-a-week children were more constricted and had a more stereotyped approach to the world. They showed a median $F+\%$ of 100 at first follow-up, and an $A\%$ that was not only significantly greater ($p \cong .06$) than that of the children seen four times a week (61 vs. 53%), but also considerably higher than the norm for this age (46%). By contrast, the $A\%$ of the children seen four times a week decreased from end to follow-up point

TABLE 4
RATE OF IMPROVEMENT IN READING, SPELLING, AND ARITHMETIC FOR CHILDREN SEEN
ONCE A WEEK AND FOUR TIMES A WEEK

Academic subject	The months before treatment	First year of treatment	Last year of treatment	First year after treatment	Second year after treatment
Reading					
Once a week	.06	.09	.08	.02	.07
Four times a week	.05	.04	.09	.11	.15
Spelling					
Once a week	.04	.06	.05	.01	—
Four times a week	.04	.03	.06	.10	—
Arithmetic					
Once a week	.06	.05	.10	.07	—
Four times a week	.06	.08	.06	.07	—

Note.—Rate of improvement was defined as gain in grade placement points divided by number of months in interval. Significance levels were determined by Mann-Whitney *U* test.

* $p \approx .057$.

** $p \approx .029$.

*** $p \approx .014$.

and a greater number of different content categories were used ($p \approx .06$).

The percentage of responses to the last three Rorschach cards has been interpreted to indicate "the general responsiveness to emotional stimuli from the environment whether this is expressed in overt reactivity or not [Klopfer et al., 1954]." This percentage was significantly lower ($p \approx .02$) for the once-a-week as opposed to the four-times-a-week children at follow-up (29 vs. 44%).

DISCUSSION

It is possible that factors other than those associated with frequency of treatment accounted for the striking differences in outcome. Important changes did occur in the impact of the family on the child, but these were no greater for one or the other group. The quality of teaching and the availability of supplemental educational experiences were also essentially equivalent for both groups. The total number of sessions was positively

associated with the various outcome indexes, but not significantly so in any instance. We would expect that if larger samples were available, the above variables would be sources of variation in their own right.

The therapist's attitude toward the various frequencies of treatment could also have accounted for the differences in outcome. Examination of the recordings of his feelings about each session revealed no differential interest in or differential expectations as to the ultimate outcome of the treatment. Moreover, if the therapist's attitude toward the efficacy of the different frequencies was the crucial factor, why did the once-a-week group initially show greater gains, and why did the two groups not differ significantly on a greater number of indexes at the end of treatment?

Examination of the therapeutic process of the two treatment situations suggested instead that frequency of session was associated with a cluster of process variables which were in turn associated with differential outcome. Striking differences in both the child's material and the therapist's ability to use that material have been found and will be reported in the future.

The quality and the pattern of changes in the outcome indexes themselves suggest that a differential personality reorganization took place as a function of the different treatment processes. It is hypothesized that the nature of the treatment of the once-a-week children was such as to allow for a more immediate

TABLE 5
MEDIAN TOTAL IQ VALUES FOR THE ONCE-A-WEEK
AND FOUR-TIMES-A-WEEK CHILDREN

Group	Point of assessment		
	Beginning of treatment	End of treatment	Year after end of treatment
Once a week	110	100	109
Four times a week	106	108	109

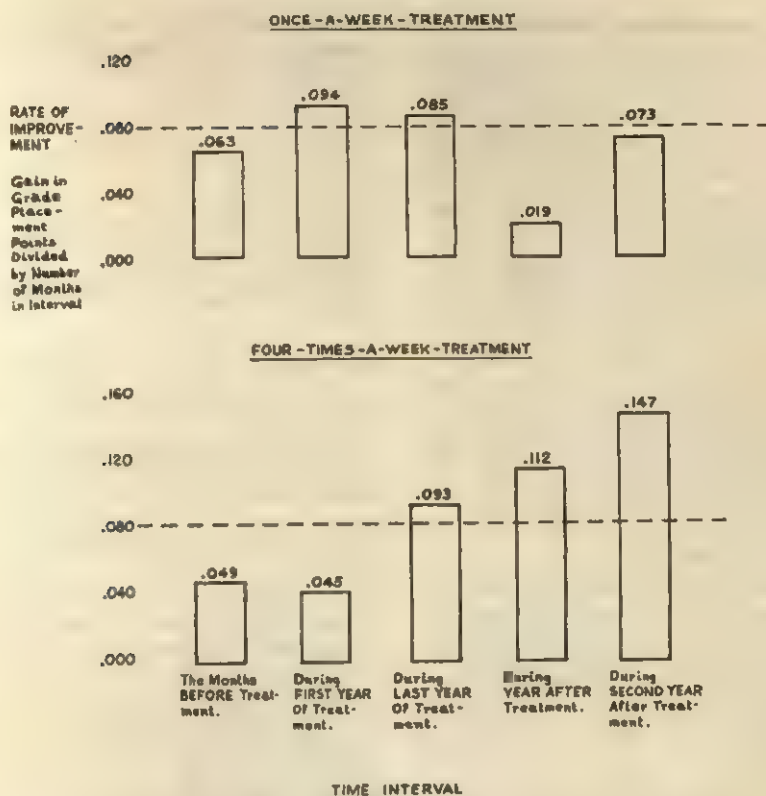


FIG. 1. Rate of improvement in reading as measured by the Wide Range Achievement Test.

improvement in the level of ego integration, and this in turn permitted more immediate advances in the academic learning situation. By contrast, the more intensive frequency facilitated the regressive repetition of all the ramifications of intrapsychic and environmental conflicts, and this would initially not favor improvement in the symptom area. But by the end of treatment it was seen that this more extensive and qualitatively different process allowed for a degree of flexibility that in combination with an adequate level of ego integration would insure continued progressive adaptation in the follow-up period (see Heinicke, 1965a, for elaboration).

The combination of outcome indexes used in this study not only facilitated the link to the process variables but has other advantages. The special knowledge of the therapist was made use of but his subjective bias was checked both by the assessments of other clinicians and by the results of the analysis

of relevant test scores. A reliable procedure for summarizing the behavioral observations of different clinicians has been developed through the use of the Diagnostic Profile and the ratings based on it. Factor analysis of these ratings resulted in meaningful factors. Scores based on one of these, namely, ego integration, were positively associated with both the absolute level and the increase in the Stanford IQ scores. This, in turn, suggests the importance of motivational factors in accounting for variations in IQ. Studies of normal populations have demonstrated a relationship between measures of IQ and ratings of personality not unlike those included in the ego integration cluster (Bayley & Schaefer, 1964; Sontag, Baker, & Nelson, 1958).

While further analysis of the present data may lead to different results and interpretations, the following general conclusions seem warranted at this time: (a) Children whose learning disturbance is a function of internal

and external psychological conflict are in many instances not likely to benefit from additional educational help; (b) psychoanalytic psychotherapy with this type of child and his family does facilitate the child's progressive development, including his use of educational opportunities; (c) the child's continuous development following the termination of treatment is greatly enhanced if he has experienced an intensive form of psychotherapy; and (d) studies of treatment which do not assess the child after termination are seriously incomplete; one of the salient characteristics of the children seen once a week was their continued dependence on the intervention and guidance of the adult to maintain their defensive organization.

REFERENCES

- BAYLEY, N., & SCHAEFER, E. Correlations of maternal and child behaviors with the development of mental abilities: Data from the Berkeley Growth study. *Monographs of the Society for Research in Child Development*, 1964, 29(6, Serial No. 97).
- EDWARDS, A. L. *Experimental design of psychological research*. New York: Holt, Rinehart & Winston, 1950.
- FREUD, A. Assessment of childhood disturbances. *Psychoanalytic Study of the Child*, 1962, 17, 149-158.
- FREUD, A. The concept of developmental lines. *Psychoanalytic Study of the Child*, 1963, 18, 245-265.
- HEINICKE, C. M. Frequency of psychotherapeutic session as a factor affecting the child's developmental status. *Psychoanalytic Study of the Child*, 1965, 20, 42-98. (a)
- HEINICKE, C. M. Notes on the strategy of a child psychotherapy project. *Reiss-Davis Clinic Bulletin*, 1965, 2, 80-86. (b)
- IMBER, S. D., FRANK, J. D., NASH, E. H., STONE, A. R., & GLIEDMAN, L. H. Improvement and amount of therapeutic contact: An alternative to the use of no treatment controls in psychotherapy. *Journal of Consulting Psychology*, 1957, 21, 309-315.
- KLOFFER, B., AINSWORTH, M., KLOFFER, W., & HOLT, R. *Developments in the Rorschach technique*. Vol. 1. Yonkers-On-Hudson, N. Y.: World Book, 1954.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LORD, F. M. The measurement of growth. *Educational and Psychological Measurement*, 1956, 16, 421-437.
- LORE, M., MCNAIR, D. M., MICHAUX, W. W., & RASKIN, H. Frequency of treatment and change in psychotherapy. *Journal of Abnormal and Social Psychology*, 1962, 64, 281-292.
- MCNAIR, D., LORE, M., YOUNG, H., ROTH, I., & BOYD, W. A three-year follow-up of psychotherapy patients. *Journal of Clinical Psychology*, 1964, 20, 258-264.
- SONTAG, L., BAKER, C., & NELSON, V. Mental growth and personality development: A longitudinal study. *Monographs of the Society for Research in Child Development*, 1958, 23(2, Serial No. 68).

(Received May 20, 1968)

A NOTE ON RELAXATION TREATMENT GROUPS IN STUDIES OF SYSTEMATIC DESENSITIZATION

WILLIAM T. McREYNOLDS¹

University of Texas

Recent criticisms by Davison and Folkins et al. of each others' relaxation training condition in the investigation of systematic desensitization are reviewed. It is concluded that a "pure" relaxation treatment condition void of their criticisms does not represent a real treatment or treatment analogue.

In a recent exchange of criticisms Davison (1969) and Folkins, Evans, Opton, and Lazarus (1969) found each other to be guilty of employing a faulty relaxation treatment group in the investigation of the comparative effects of systematic desensitization (Wolpe, 1958), relaxation therapy (Jacobson, 1938), and "extinction" or "cognitive rehearsal" (i.e., imagining disturbing situations without relaxation training). Davison's contention is that the relaxation condition employed by Folkins, Lawson, Opton, and Lazarus (1968) was vitiated by the latters' instructions to Ss that "relaxation . . . can be used to your advantage to handle the somewhat stressful movie film you are about to see. Just remain as relaxed as you can [p. 103]." This procedure of instructing Ss to relax *during* the presentation of actual aversive stimulus material was seen by Davison to change the nature of the "treatment" condition such that it more closely represented "in vivo desensitization" than relaxation alone. Davison (1969) feels this is true because "Folkins et al.'s 'relaxation group' seems to have experienced the apparently crucial *pairing* of anxiety-competing relaxation with stressful stimuli [p. 87]."

The retort by Folkins et al. contended that it is Davison, himself, who has employed a faulty relaxation group in his study of systematic desensitization as a "counter conditioning" process (Davison, 1968). This, they reason, is true because Ss undergoing relaxation training in Davison's study knew at the

time of their treatment the purpose of the treatment (to reduce fear of snakes) and, thus, "had ample opportunity and motivation to think about snakes during the relaxation training [pairing anxiety-competing relaxation with the disturbing stimulus] and hence to contaminate the relaxation treatment with elements of desensitization [Folkins et al., 1969, p. 88]." Further, Folkins et al. argue that their relaxation training procedures avoided this "flaw" since Ss in their study knew only that they were to see a stressful film but were unaware of the content of that film until the actual showing. Folkins et al. concluded that desensitization could have very little opportunity to occur, and even though their Ss were explicitly told to make use of relaxation training during the stressful situation while Davison's were not, the procedures in the latter study implicitly encouraged Ss to do so.

What both Davison and Folkins et al. appear to be advocating for the study of the separate effects of relaxation and conditioning in systematic desensitization is a "pure" relaxation treatment condition (i.e., one free of desensitization effects) in which Ss are not aware of the target behavior (e.g., reaction to a film of a fatal accident, fear of snakes). The present author seriously questions such a treatment condition. Unless Ss know the focus of treatment and the effect treatment is expected to have on them, the nonspecific placebo effects of their expectations and aspirations for change will not contribute to the final outcome. This point is especially crucial in comparisons between Ss receiving relaxation treatment and Ss undergoing systematic desensitization (or any other form of treatment) since, as Folkins et al. (1969) pointed out, Ss in desensitization will, by virtue of the

¹The author thanks Janet T. Spence of the University of Texas for her suggestions and comments in the preparation of this paper.

Requests for reprints should be sent to the author who is now at the Department of Psychology, University of Kentucky, Lexington, Kentucky 40506.

nature of the pre- and posttreatment tests and the treatment procedures (e.g., anxiety hierarchy), have a good idea of the goal and purported effects of treatment. The omission of these expectation and other placebo effects will have an undefinable and possibly serious, detrimental effect on the effectiveness of treatment (Frank, 1959, 1961; Imber, Frank, Nash, Stone, & Gliedman, 1957). Without the knowledge of the goal and expected result of treatment, the experience of relaxation cannot be expected to have more relevance to Ss' "symptomatic" behavior than any other pleasant experience.

Davison further contended that for a relaxation condition to be free of flaw Ss must not be instructed to attempt to maintain relaxation during crucial disturbing situations. Stated differently, Davison's contention is that there must be no "application" instructions if the relaxation treatment is to be "relaxation alone." Folkins et al. (1969) noted that the inclusion of such instructions makes the relaxation training more similar to the clinical application of treatment by relaxation. The present author would go one step further and insist that instructions to use relaxation in the face of disturbing situations are a necessary and important part of relaxation treatment. Jacobson (1938) stated: "The aim is to train the patients to use *his own initiative*. He learns to localize tensions when they occur during nervous irritability and excitement and to relax them away [p. 40]." This aspect of relaxation training is often overlooked by researchers and may prove to be a crucial variable in relaxation treatments² (Zeisset, 1968).

² W. T. McReynolds, Systematic desensitization, insight-oriented psychotherapy, and relaxation therapy in a psychiatric population. Unpublished doctoral dissertation in preparation, University of Texas, 1969.

In the light of the preceding discussion it appears that the weaknesses Davison (1969) and Folkins et al. (1969) find in each others relaxation condition are in fact their real strengths, and vice versa. It may be that the resolution of their discrepant results will be found in other, less subtle procedural differences than the ones pointed out by the authors themselves.

REFERENCES

- DAVISON, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, 73, 91-99.
- DAVISON, G. C. A procedural critique of "Desensitization and the experimental reduction of threat." *Journal of Abnormal Psychology*, 1969, 74, 86-87.
- FOLKINS, C. H., EVANS, K. L., OPTON, E. M., JR., & LAZARUS, R. S. A reply to Davison's critique. *Journal of Abnormal Psychology*, 1969, 74, 88-89.
- FOLKINS, C. H., LAWSON, K. D., OPTON, E. M., JR., & LAZARUS, R. S. Desensitization and the experimental reduction of threat. *Journal of Abnormal Psychology*, 1968, 73, 100-113.
- FRANK, J. D. Problems of controls in psychotherapy as exemplified by the psychotherapy research project of the Phipps Psychiatric Clinic. In E. A. Rubenstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- FRANK, J. D. *Persuasion and healing*. Baltimore: Johns Hopkins University Press, 1961.
- IMBER, S. D., FRANK, J. D., NASH, E. H., STONE, A. R., & GLIEDMAN, L. H. Improvement and amount of therapeutic contact: An alternative to the use of no-treatment controls in psychotherapy. *Journal of Consulting Psychology* 1957, 21, 309-315.
- JACOBSON, E. *Progressive relaxation*. Chicago: University of Chicago Press, 1938.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.
- ZEISSET, R. M. Desensitization and relaxation in the modification of psychiatric patients interview behavior. *Journal of Abnormal Psychology*, 1968, 73, 18-24.

(Early publication received May 10, 1969)

SEX-ROLE STRAIN AND PERSONALITY ADJUSTMENT OF CHINA-BORN STUDENTS IN AMERICA:

A PILOT STUDY¹

STANLEY L. M. FONG

AND

HARVEY PESKIN²

State University of New York at Buffalo

San Francisco State College

Since the role of student has been traditionally reserved for the male in Chinese society, it was hypothesized that China-born, *female* college students would feel generally alienated from and in defiance of norms of the parent culture. Results on the California Psychological Inventory (CPI) with 43 male and 43 female China-born college students supported this view, with females scoring lower on Socialization and Good Impression. The Chinese model of the feminine role (reserve, patience, modesty, etc.) was most rejected by the subsample of females on student visas, and most accepted by the naturalized females (including permanent residents), as reflected on the CPI Femininity Scale. Regardless of sex, naturalized and resident students clearly outranked visa students on important aspects of psychological health.

People migrating to another society can be expected to carry over the culturally ingrained emotional meanings of essential social roles. When two cultures are as different as the Chinese and the American, new role meanings often come into conflict with traditional ones, with evident implications for personality adjustment under the new circumstances. Moreover, such discrepancies may be greater for one sex than the other, and thus pose interesting questions relating to sex roles. This study is concerned with the differential adjustment of China-born male and female students in America.

In the role of student, the sexes doubtless face differential role strain in Chinese culture. Since the sixth century, males only, rich or poor, were allowed to take the imperial examinations, based on the Confucian teachings (Hsu, 1955). In 1905, the imperial examinations were abolished before the decline of the Manchu dynasty and the establishment of the people's republic. But the traditional respect for the male's scholarship is still very

much in evidence, with the college degree replacing the imperial ones. Education for the woman was seen to endanger feminine virtue. Her status in life was to be a devoted helper to the future mother-in-law and a bearer of many grandchildren for her, preferably male ones. It was only in the last century, with the aid of missionaries, that the emancipation of Chinese women began, and a number of Christian schools and colleges were established. As late as 1910, these schools alone provided women with the same educational opportunities as men (Liu, 1963). Although their educational ambitions have enlarged, Chinese girls may still find that their emerging role is not fully accepted by all segments of Chinese society (DeVos & Abbott, 1966; Wright, 1964).

From this brief analysis, it is reasonable to suppose that females in college will feel alienated from and rebel against traditional expectations of what is appropriate for Chinese females.

A host of sociopsychological factors may either mitigate or aggravate such sex differences. Certainly, length of residence in America would be expected to facilitate the incorporation of the new culture. Over and above residence, the Chinese's eligibility for naturalization, under liberalized immigration policy, has conferred the constitutional and psychological security of full membership in

¹ This study is based on a paper presented at the Western Psychological Association Convention, San Francisco, 1967.

² Also Research Associate, Institute of Human Development, University of California, Berkeley. Requests for reprints should be sent to Harvey Peskin, Department of Psychology, San Francisco State College, 1600 Holloway Avenue, San Francisco, California 94132.

TABLE 1
MEAN SCORES OF NATURALIZED AND VISA MALES AND FEMALES ON THE CALIFORNIA
PSYCHOLOGICAL INVENTORY

CPI scales	Males		Females	
	Visa (N = 23)	Naturalized (N = 20)	Visa (N = 22)	Naturalized (N = 21)
	M	M	M	M
Dominance	47.7	50.9	46.8	51.0
Capacity for Status	44.3	49.6	42.3	43.5
Sociability	45.9	49.0	41.2	47.7*
Social Presence	44.6	51.0*	44.5	49.8
Self-Acceptance	48.9	52.8	46.4	52.1
Sense of Well-Being	39.1	46.5*	35.5	40.7
Responsibility	45.2	48.8	41.6	48.2**
Socialization	50.3	52.9	46.2	46.8
Self-Control	46.4	49.1	46.3	43.1
Tolerance	40.0	48.3**	40.1	44.9
Good Impression	47.3	49.6	44.6	42.1
Communality	45.3	52.5**	45.0	48.3
Achievement via Conformance	45.2	51.7**	43.5	46.7
Achievement via Independence	45.9	54.1*	47.2	51.1
Intellectual Efficiency	40.1	46.1	39.5	43.1
Psychological Mindedness	49.2	50.0	46.4	52.2*
Flexibility	44.6	51.0*	49.3	53.6
Femininity	55.7	55.1	50.0	58.6***

* $p < .10$.

** $p < .05$.

*** $p < .01$.

American society. Thus, it may be expected that the naturalized Chinese (or those eligible for citizenship) will be noticeably more in tune with significant American norms and thus with themselves, in terms of well-being and self-esteem, than a disenfranchised Chinese group. For this study, the latter will be represented by Chinese holding student visas.

METHOD

Subjects

The Chinese Ss were students, living and attending colleges in the San Francisco Bay area. They were all born in China, and the sample consisted of 43 males and 43 females. Of these Ss, 20 males and 21 females were naturalized American citizens or permanent residents eligible for citizenship. (Seven of these 41 Ss fell into the latter; these 7 have only recently been able to satisfy the residency requirement for citizenship and presumably plan to file for citizenship in the near future.) These 41 Ss are referred to as the naturalized male and female groups in this study. Twenty-three males and 22 females were on student-visa status. The mean ages of these four groups of male and female Ss were 22.3, 21.2, 24.0, and 20.8 yr., respectively. The naturalized Ss have lived in the United States for 11.5 mean yr., and the visa Ss, 3.5 yr. Most of the Ss were born on the China mainland (65%); the remainder came

from Hong Kong (30%) and Taiwan (5%). Generally, the Chinese from the China mainland migrated before the establishment of the Communist government.

Procedure

Most of the Ss were obtained from psychology courses at San Francisco State College. Since the students are required to take two courses in psychology to meet the requirements of the college, the present sample represents a wide spectrum of curriculum majors. A few students at other colleges were also given the questionnaires. The questionnaires were filled out anonymously by Ss and returned by mail. Sixty-five percent of the questionnaires were returned. It is hoped that later studies will obtain a better rate of return than has been possible in this voluntary study.

All of the Ss were given a battery of questionnaires which included a Personal Data Form (Fong, 1965) and the California Psychological Inventory (CPI, Gough, 1957). The present report considered only the CPI whose relevance for cross-cultural research has been repeatedly demonstrated (e.g., Gough, 1960, 1966; Gough & Sandu, 1964).

RESULTS

The mean scores of the visa and naturalized males and females on the CPI scales are presented in Table 1. A two-way analysis of vari-

ance computed by the method of least squares for unequal cell frequencies (Winer, 1962) was performed on each scale with main effects of sex and citizen status. Results are presented in Table 2.

The effect of sex yielded significant differences on two scales, Socialization and Good Impression, with the males scoring higher on both.

The effect of citizen status (i.e., visa or naturalized) yielded significant differences on 10 of the 18 CPI scales. The naturalized Ss obtained higher scores than the visa Ss on all of them: Social Presence, Sense of Well-being, Responsibility, Tolerance, Communality, Achievement via Conformance, Achievement via Independence, Intellectual Efficiency, Flexibility, and Femininity. Since the last named scale owed its significance entirely to the wide gap between the female groups (see Table 1), it is more accurate to speak of a yield of 9 significant scales on the main effect of citizen status. The Femininity scale was the only one which showed a significant interaction between sex and citizen status. The naturalized females scored significantly higher on Femininity than the visa females.

DISCUSSION

Males and Females

It was assumed that the female students would experience greater conflict with and rebellion against social norms and role expectations than the male students. This prediction is supported by the lower scores of females on Socialization and Good Impression, the two sex-discriminating scales of this study. Studies (e.g., Gough, 1965; Stein, Gough, & Sarbin, 1966) have shown that the Socialization scale deals basically with role-taking disposition and relates to deviant or rule-breaking behavior. The low scores of the Chinese females on Good Impression indicate that they either reject or are less aware of the socially desirable response to enhance self-esteem in their social groups. The present view would, of course, presume that such scores reflect the same deliberate disavowal of cultural role demands that had been operating in their decision to attend college, and possibly to pursue a career.

TABLE 2
ANALYSIS OF VARIANCE OF CPI SCALES

Scale	Sex (S)	Citizen status (C)	S × C	Within
Social Presence				
MS	9.67	733.33	6.30	107.94
F	.10	6.79*	.06	
Responsibility				
MS	100.77	561.69	50.34	83.32
F	1.21	6.74*	.60	
Socialization				
MS	555.05	50.93	21.82	87.18
F	6.37*	.58	.25	
Tolerance				
MS	53.03	920.18	68.59	109.51
F	.48	8.40**	.63	
Sense of Well-being				
MS	470.77	841.79	26.01	149.71
F	3.15	5.62*	.17	
Good Impression				
MS	526.74	.23	122.95	75.04
F	7.02*	.00	1.64	
Communality				
MS	94.57	582.11	83.59	98.13
F	.96	5.93*	.85	
Achievement via Conformance				
MS	220.74	497.91	58.63	77.75
F	2.84	6.40*	.75	
Achievement via Independence				
MS	10.47	777.50	95.52	104.42
F	.10	7.45**	.92	
Intellectual Efficiency				
MS	63.34	499.26	30.11	107.95
F	.59	4.63*	.28	
Flexibility				
MS	296.42	601.17	23.37	121.49
F	2.44	4.95*	.19	
Femininity				
MS	35.15	342.51	460.11	85.33
F	.41	4.01*	5.39*	

Note.—*df* = 1 for sex, citizen status, and S × C; *df* = 82 for within.

* *p* < .05.

** *p* < .01.

Visa and Naturalized Status

The psychological consequences of the second variable, naturalization, have not, as far as the authors know, been reported before. Yet its effects seem to be large indeed. The naturalized Ss presented a picture of being more self-confident (higher Social Presence), freer from self-doubt and disillusionment (higher Sense of Well-being), more responsible and dependable (higher Responsibility), more accepting and tolerant of others (higher Tolerance), more flexible and adaptive in their thinking and social behavior (higher

Flexibility), more productive in their intellectual functioning (higher Achievement via Conformance, higher Achievement via Independence, higher Intellectual Efficiency), and more in touch with the implicit and underlying values of the American scene (higher Community). The visa Chinese, on the other hand, seemed to experience a degree of discomfort severe enough to disorder their functioning in these psychological and social areas.

Understandably, immigration status and length of American residence were confounded in this study. Whether length of residence, irrespective of citizenship (or eligibility), contributed to the higher CPI scores of the naturalized group can be gleaned from within-groups correlations. For all of these nine significant CPI scales, correlations within the naturalized group were positive, but only significantly so on the Responsibility scale ($p < .10$), that is, those residing longer in the United States scored higher on this measure. It is reasonable to suggest, therefore, that citizenship, or the prospect of citizenship, encourages a motivational state in which personal commitment to the new society is achieved by the full granting of its legal guarantees, relatively independent of the actual time of residence.

It is instructive to inspect the pattern of sex differences among the nine scales that showed a significant main effect for citizen status. (The Femininity scale is not considered here.) On eight of the nine scales, significant t 's ($p < .10$) were obtained between the visa and naturalized groups within one or the other sex, but not within both sexes (Table 1). On seven of the eight scales, it was the difference between the male groups of visa and naturalized students which was significant; only on one scale did the female groups differ. Evidently, the contribution of the male groups to the main effect of citizen status outweighed that of the female groups. The sharper separation of the male groups may reflect the greater social disruption experienced by Chinese men on student-visa status. Being officially consigned to the limbo of "student-hood" while in America, without the legal right to plan a professional career here and

with the considerably fewer employment opportunities in Free China, may constitute a blow to one's identity as a man, especially to those in their midtwenties.

Femininity

The Femininity scale is of particular interest on two counts. First, its cross-cultural validity has been perhaps the best established of the CPI scales (Gough, 1966). Second, in this study, it is the only scale on which the Chinese males and females exceeded the American norms. This finding accurately mirrors the Chinese cultural emphasis on gentleness, modesty, patience, reserve, and social sensitivity. In the present study, it was the naturalized female students who scored the highest and the visa female students who scored lowest, with the male groups falling in between. That the visa female students, in terms of the overall Chinese average, should be least feminine is not surprising in view of the unusual self-reliance and independence which the Chinese woman must show to make her way into higher education. Considered together with their low scores on Socialization and Good Impression, the visa females, as a special subgroup, would then appear to follow a style of the "masculine protest," so well known in Western characterology (cf., Yuntang's, 1935, protest against "those masculine women prize winners at college.") But a surprisingly different style seems to characterize the naturalized females. Whereas their equally low scores on Socialization and Good Impression indicate a similar social estrangement, their high Femininity scores point to a compensatory compliance with feminine role demands. Since almost all live in America with both parents (20 of the 21 girls compared to 7 of the 22 visa girls), parental influence is more assured, not merely for role modelling, but as a force to be appeased. We might say that the naturalized females are in an inescapable dilemma, from which the visa females are at least spared: that of making—or pretending—peace between their educational ambitions in the permissive atmosphere of American coeducation and the antifeminist values of Chinese parents with whom they continue to live in America.

This dilemma may diminish where older siblings had already helped to accommodate the family to western values. The probable influence of older siblings on acquiring the feminine role is suggested by the diminishing Femininity mean scores of the naturalized female firstborns (65.5), second borns (58.0), and later borns (53.3). Firstborns exceed the latter groups, $p < .05$. (For the visa females, all birth-order groups are essentially equivalent.) Interestingly, no similar birth-order differences for the naturalized females occur on the Socialization and Good Impression scales, where mean scores are essentially equivalent across birth order. We are perhaps left to conclude that the loyalty of naturalized firstborns to Old World norms of femininity is more apparent than real, that is, more in the nature of role playing than an authentic acceptance of the Chinese status quo for women. It may also follow that these Ss' parents (their own citizenship status notwithstanding) are seen by their daughters as carriers of the ethnic culture who must be mollified. Perhaps particularly toward sex roles, where parental attitudes are likely to have been settled for good and all, a display of obedience by the first offspring may both disguise and betray the alienation between the generations.

REFERENCES

DEVOS, G. A., & ABBOTT, K. A. (Advisors) The Chinese family in San Francisco: A preliminary

- study. Unpublished group masters' thesis. University of California, Berkeley, 1966.
- FONG, S. L. M. Assimilation of Chinese in America: Changes in orientation and social perception. *American Journal of Sociology*, 1965, 71, 265-273.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- GOUGH, H. G. Theory and measurement of socialization. *Journal of Consulting Psychology*, 1960, 24, 23-30.
- GOUGH, H. G. Cross-cultural validation of a measure of asocial behavior. *Psychological Reports*, 1965, 17, 379-387.
- GOUGH, H. G. A cross-cultural analysis of the CPI femininity scale. *Journal of Consulting Psychology*, 1966, 30, 136-141.
- GOUGH, H. G., & SANDHU, H. S. Validation of the CPI socialization scale in India. *Journal of Abnormal and Social Psychology*, 1964, 68, 544-547.
- Hsu, F. L. K. *Americans and Chinese*. London: Cresset, 1955.
- LIU, K. C. *Americans and Chinese*. Cambridge, Mass.: Harvard University Press, 1963.
- STEIN, K. B., GOUGH, H. G., & SARBIN, T. R. The dimensionality of the CPI socialization scale and an empirically derived typology among delinquent boys. *Multivariate Behavioral Research*, 1966, 1, 197-208.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WRIGHT, B. R. Social aspects of change in the Chinese family pattern in Hong Kong. *Journal of Social Psychology*, 1964, 63, 31-39.
- YUNTANG, L. *My country and my people*. New York: Reynal & Hitchcock, 1935.

(Received April 22, 1968)

PREDICTION OF REACTION TO PLACEBO¹

MICHAEL M. NASH AND FRED M. ZIMRING²

University of Chicago

Expectations of drug effect and openness to experience were employed as predictors of placebo response. Ninety-nine Ss (mean age 82) received either drug or placebo for 8 days. Before drug therapy S's short-term memory was tested, expectations about the pill determined, and openness to experience ascertained. After therapy memory was tested and S questioned about change experienced. An analysis of variance revealed that the two experimental drugs had no noticeable effect. A similar analysis of expectations yielded significant *F* ratios. Openness had no predictive value. Expectations correlated significantly with the subjective measure of placebo reaction (experienced change, $p < .001$) as well as with a behavioral criterion (improved short-term memory, $p < .03$).

Within the last decade there has been increasing interest in drug research with human Ss. A vast range of drugs has been tested including compounds as different in their behavioral effects as tranquilizers and hallucinogens. Personality research has also been intense, especially the study of individual differences.

Despite this activity, a particular topic which overlaps the two fields has been consistently difficult to explore. Uhr (1960) pointed out that one of the rarest of objective drug experiments is the study of individual personality differences that correlate with a person's response to drug therapy. Six years later the problem was the subject of a special workshop on *Prediction of Response to Pharmacotherapy* (Wittenborn & May, 1966). In the published report of the proceedings, however, the only chapter which dealt specifically with the actual prediction of drug-induced changes was concerned with avoidance behavior in rats. Most discussions of individual differences and drug effect have been limited to post facto attempts at psychodynamic analysis (Kubie, 1960). The few studies which separated the influence of drug from personality have used mentally ill populations and tranquilizing drugs (Kurland,

1960; Wittenborn & May, 1966). Closely related to the problem of the influence of personality on drug action is the phenomena of an experienced or observed drug effect when the substance ingested is actually inert. Such substances, called placebos, are agents which although pharmacologically impotent are employed with the definite suggestion that they will be helpful. Any changes which occur after ingesting a placebo ("placebo effect") are interpreted as being primarily psychological in origin.

The manner in which suggestion can interact with drugs, especially in producing the placebo effect has been of great theoretical interest (Katz, 1967). The approach most psychopharmacologists would prefer would be to predict who will react to a placebo in a given drug study in order to control for that effect. That is, if it would be possible to assess the degree to which an S possessing a particular trait would react to a placebo, then this effect could be taken account of when assessing the effect of the actual drug being tested. Making predictions of this nature, however, on the basis of previously measured personality characteristics has thus far met with almost no success (Peterson, 1966).

The present study was an investigation of the relationship between two carefully selected personality characteristics and susceptibility to two drugs and to a placebo; the drugs and the placebo presented to Ss as having the effect of improving short-term memory.

¹ This study was supported by the Dreyfus Fund. Paul Gordon and Sheldon Tobin were the principal investigators.

² Requests for reprints should be sent to Fred M. Zimring, Counseling and Psychotherapy Research Center, 5846 University Avenue, University of Chicago, Chicago, Illinois 60637.

The immediate goal of the study was to predict which *Ss* would react to the placebo.

The two personality characteristics selected as independent variables were: (a) *S's* expectations (EXP) toward the drug he was to receive and (b) *S's* openness (OPN) to new experience. Susceptibility, the dependent variable, was defined as the degree to which *S's* performance improved on three tasks emphasizing short-term memory (STM) and the amount of positive experienced change (ECH) he reported. The drugs used were diphenylhydantoin (DPH), procaineamide (PRO), and a placebo (PLA).

The first of these personality variables, *S's* EXP, is viewed as an important component of *S's* attitude. Several investigators (Frank, 1961; Katz, 1967; Kubie, 1960) have felt that *S's* attitude is perhaps the most central factor in any therapy program in which the concern is changing *S's* behavior. Optimum change appears to require a positive, trusting attitude toward the agent of change. In drug therapy these agents are the *Es* as well as the drugs. In the present study, *S* was told that there was a drug which might help him remember better, was assured it would not harm him, and was asked to volunteer. The degree to which *S* trusted the *Es* and actually expected to be helped was called *S's* EXP and was construed as being important to the placebo response.

Psychotherapy research by Rogers and others (Rogers & Dymond, 1954) has shown that a client's OPN to new experience is related to positive change while in therapy. In the present study, OPN is taken to mean the ability to be aware of, and to be articulate about, events occurring within oneself. It seemed likely that persons who rate high on OPN would show reaction to placebo. Such open *Ss*, sensitive to any variance within themselves, were thought to be likely to believe that the small random changes in their daily physical and emotional feelings were drug induced.

The drugs used in the study belong to a class of anticonvulsants which act to raise the threshold for neuron firing. It is suspected (Gordon, 1967) that a decreased threshold for neuron firing with increasing age interferes with STM.

Both DPH and PRO have been shown to improve the immediate memory of certain strain of aged rats (Gordon, Tobin, Doty, & Nash, 1968). The placebo reactor study was designed as part of a larger research project comparing the effects of the two drugs to that of the placebo. When it was determined that DPH and PRO had not affected *Ss* differently from PLA on any dimension, both drugs were then viewed as placebos and the entire sample handled as three groups given placebos. Thus the results reported occurred within both drug-as-placebo groups and the actual placebo group.

METHOD

Subjects

A potential population of 200 *Ss* was drawn from the residents of a Jewish home for the aged in Chicago, Illinois. In a series of meetings *Ss* were told that a new pill had been developed which might improve their memory for recent events. They were assured the pill would not harm them and were offered a small sum of money if they would volunteer to take the pill for a week and answer some questions. One hundred and one residents joined the program, 2 withdrew before receiving the drugs, leaving a final sample of 99. Three volunteers had to be refused because of other medication they were receiving. The *Ss* ranged in age 62-94 with a mean age of 82. Twenty-two men and 77 women participated.

Procedure

The sample was divided on a random basis into three groups and 34 *Ss* assigned DPH, 32 PRO, and 33 PLA. A double blind design was used. In this type of design, drug identity is coded and is unknown to the nurses, interviewers, *Ss*, and *Es*. In addition to this procedure, all outcome variables were scored and analyzed without knowledge of the corresponding predictive measures. Experienced nurses administered the pills and kept close watch for side effects. Four trained interviewers tested a specific group of *Ss* through the entire study.

Each *S* was told by his interviewer that the purpose of the study was to see if the drug would affect their memory and tested to establish his initial level of STM ability and to obtain EXP and OPN measures. This interview, designated Time 1 (T_1), was conducted 1 wk. prior to *S* receiving one of the drugs. The second interview (T_2) took place about 3 hr. after *S* had received his first drug administration. This interview was included to test the reliability of some instruments and to check for possible acute drug effect. After *S* received either DPH, PRO, or PLA each morning for 8 days, he was tested a third and final time (T_3). Data from T_3 included measures of STM change and experienced change

TABLE 1
CORRELATIONS OF PREDICTIVE VARIABLES
WITH RESPONSE TO PLACEBO

Hypothesis	R
OPN T ₁ with experienced change T ₂	.097
EXP T ₁ with experienced change T ₂	.355**
EXP T ₁ with short-term-memory improvement T ₁ to T ₂	.231*

Note.—N = 99; OPN = openness; EXP = expectations;
T = time.
* $p < .03$.
** $p < .001$.

(ECH). Following the conclusion of the testing, data from T₂, T₃, and T₄ were scored and the drug code revealed. Each group was then compared and found to be equivalent in age, sex, and at T₁ to be equal in initial level of EXP, OPN, and STM.

Measures

OPN was measured at T₁ by asking S for stories for Cards 1 and 2 of the Thematic Apperception Test (TAT) as well as for a third card especially designed for geriatric populations called Park Bench (an old man and an old woman sitting together on a bench in a park). Each story was scored for OPN by using an affect-cognition scale developed by Rachman (1965) based on work by Zimring (1962, 1963). From 1 to 4 points was given for a response depending on the degree to which affect was merely mentioned, attributed to a character, or developed within a conceptual framework. Subscores for the three cards were then totaled and the sum designated as OPN to experience.

EXP was evaluated at T₁ by use of a questionnaire asking about S's belief in the Es and in the drug and also asking about the ways and amount S hoped the drug would help him. Eight questions were each scored on a 3-point scale, 2 points being given if S indicated that he had positive feelings and expectations, 1 point if a neutral or indifferent answer was given, and 0 points if S was pessimistic or cynical in his reply. The S's scores on all eight questions were then summed and recorded as his EXP. For example, the median score of 7 or 8 on this variable depicted an S who was either noncommittal or unconcerned about what he felt the drug's effects might be.

In addition, at this time the interviewers were asked to fill out a status sheet which required them

to make judgments (on a 5-point scale) about 12 personality characteristics of S. For example, S was rated on how cooperative, suspicious, and anxious he was during an interview.

ECH was measured at T₁ by a set of eight questions intended to force S into assessing the particular ways in which he experienced the drug as having helped (for example, did he think his memory had improved?) as well as eliciting his more global feeling about his experience (would he recommend the pill to his friends?). From 0 to 3 points were given for a response depending on the degree to which it indicated a positive reaction to pharmacotherapy and the scores were totaled across the eight questions. The median score of about 8 on this variable was typical of an individual who reported no feeling of change, either positive or negative, as a result of medication. The experiences reported, however, ranged from very enthusiastic to totally critical.

Predrug STM ability was obtained from S at T₁. This was done by combining his scores on modified versions of Digit Span, Logical Memory and Mental Control from the memory scale developed and standardized by Wechsler (1945). The same three tests were given at T₂. Improvement was defined as fewer errors made on the tests at T₂ than at T₁. The improvement seen in STM scores from T₁ to T₂ for those Ss with high EXP was not due to the effects of initial level of STM ability. This level was the same for both high and low EXP groups.

RESULTS

An analysis of variance of the drugs (see Table 3) revealed that DPH and PRO did not differ significantly from PLA on any of a dozen outcome variables including ECH and STM. Since a drug effect had not been demonstrated, all three groups were combined and treated as a placebo group to simplify interpretation of the results.

Hypothesis 1

The greater S's OPN prior to receiving a drug (T₁), the more positive is his ECH at T₂.

Table 1 reveals that this hypothesis was not supported by the data. Predrug OPN was not

TABLE 2
DISTRIBUTION OF PREDICTIVE VARIABLES

Variable at T ₁	Lower third of scale	Middle third of scale	Upper third of scale	Range of scores	Mdn	Skew
OPN	52%	28%	16%	1-13	4	.568
EXP	34%	33%	33%	1-15	7	.2

Note.—T = time; OPN = openness; EXP = expectations.

TABLE 3
ANALYSIS OF VARIANCE: DRUG ALONE

Outcome measure	Diphenylhydantoin	Procaineamide	Placebo	df	F	p
Short-term—memory change						
Raw score	1.2	.09	4.24	2/90	1.4	.24
Scaled score	.30	.28	.50	2/90	.18	.83
Experienced change						
Raw score ^a	9.57	9.03	8.0	2/90	1.58	.21
Scaled score ^b	.45	.31	—0	2/90	1.06	.35

^a One-way analysis by expectations alone, $F = 6.54$, $p = .01$.

^b One-way analysis by expectations alone, $F = .800$, $p = .006$.

a good indicator of posttreatment ECH. Although the correlation was in the predicted direction it was far from significant. Test-retest (T_1 – T_2) reliability on the OPN measure was .65, a reflection of both the performance variability inherent in a geriatric population and the nature of the TAT itself. However, the most probable reason for failure to find the predicted relationship is that the distribution of scores on this measure fell disproportionately toward the low end of the scale as seen in Table 2. Because S 's T_1 OPN score was not known until after T_2 , the E s were unable to control for the skewed distribution. There were so few S s who rated high on OPN that either the measure used is unsuitable for aged S s, or such S s differ markedly as a group on this variable from the college students on whom the instrument was developed, and who gave an approximately normal distribution of scores. No conclusions as to the predictive value of the OPN variable for placebo reactors can be made at this time.

Hypothesis 2

The higher the EXP S has at T_1 , the more positive is his ECH at T_2 .

This hypothesis was clearly supported by the data (see Table 1). Those people with the greatest pretherapy expectations reported experiencing the most positive change. The EXP scores, when split into high, moderate, and low EXP groups and subjected to an analysis of variance yielded significant F ratios regardless of whether S had received DPH, PRO, or PLA (Table 3). This indicates that it was high EXP and not drug action which brought improvement in some S s.

Hypothesis 3

The greater the EXP S has prior to taking the drug, the greater is his improvement on STM tasks.

This hypothesis was also clearly supported by the data as shown in Table 1. Those individuals with the highest expectations were indeed the strongest placebo reactors on an objective measure of change (STM). The correlation is lower than for ECH but it should be noted that the STM change represents a specific, observable change in ability in reaction to the placebo, rather than a report of subjective experience.

No significant relationships between any of the experimental variables (OPN, EXP, STM, ECH) and age, sex, or initial level of STM ability were found. Table 4, the intercorrelations between the four experimental variables, indicates that the only significant relationships between EXP and ECH and EXP and STM change. The predictive variables, OPN and EXP, were essentially independent as were the two outcome measures ECH and STM. In addition, data from the placebo reactors (those S s with high EXP) were analyzed for an acquiescence response set.

TABLE 4
INTERCORRELATIONS OF EXPERIMENTAL VARIABLES

Variable	OPN	EXP	ECH
OPN T_1			
EXP T_1	.180		
ECH T_2	.097	.355**	
STM T_2	.117	.231*	— .044

Note.—OPN=openness; EXP=expectations; ECH=experienced change; T=time; STM=short-term-memory change.

* $p < .05$.

** $p < .01$.

TABLE 5

SUBJECTS SHOWING CHANGE, TIMES 1-3, ON EXPERIENCED CHANGE AND SHORT-TERM MEMORY

EXP	Both ECH and STM positive	No change or split scores	Both ECH and STM negative
High (N = 34)	22	11	1
Moderate (N = 31)	13	14	4
Low (N = 34)	5	14	15

Note.—EXP = expectations; ECH = experienced change; STM = short-term memory.

If mere acquiescence had been demonstrated, the *Es* could have expected to find that “yes” saying on both EXP and ECH had resulted in a significant correlation (high EXP and positive ECH). Although this would explain the finding on Hypothesis 2, there would not be any reason to expect that acquiescence had resulted in improved STM for this group. Table 5 shows clearly that more *Ss* with high EXP had positive STM as well as ECH scores than *Ss* with only moderate or low EXP. This consistency in behavioral as well as subjective improvement among the high EXP group eliminates acquiescence as a reasonable explanation of the placebo response.

DISCUSSION

Although the *Es*’ as well as the *Ss*’ expectations were for the most part confirmed, several limitations of the study should be noted.

TABLE 6

CORRELATIONS OF PERSONALITY VARIABLES WITH EXPECTATIONS

Variables	R at Time 1
Positive	
Spontaneous	.169
Friendly	.276**
Rapport	.341**
Cooperative	.351**
Happy	.390**
Motivated	.515**
Total positive	.445**
Negative	
Impulsive	.047
Suspicious	-.081
Need for support	-.109
Careful	.195
Anxious	-.220*
Confused	-.392**
Total negative	-.288**

* $p < .05$.
** $p < .001$.

The measures of EXP and ECH were developed for this study and should be validated by using them with other *Ss* receiving other drugs. Caution must always be exhibited in generalizing from one sample to another: in particular the age of the *Ss* in this study may limit the usefulness of the findings. Given the fact, however, that the aged person is almost twice as likely to *require* drugs as would a person in his 20s, the use of geriatric populations in drug research (as compared, for example, to college students) is advisable at least in terms of implications for clinical practice.

A possible explanation for improvement in STM seen in the placebo (high EXP) group is that high EXP meant that the aged *S* was sufficiently mobilized to take better advantage of the practice effects involved in having his memory tested three times. This very tentative view should be subjected to further investigation. It is interesting to note that although all 99 *Ss* volunteered for the study a full third of them (those with low EXP) had very little belief that the pill would help them and apparently were willing to take the drug for other reasons than its intended purpose.

The *Es* had hoped to uncover additional personality variables besides EXP and OPN which might be useful in identifying the placebo reactor and gaining some insight into their dynamics. Table 6 is a list of other variables, clinical assessments of *S* by the interviewer at T₁, which correlated with EXP. Hopefully some of these variables will have predictive value in future studies of the placebo response. If this psychological effect can be understood it would be possible to accurately assess the pharmacological value of the therapeutic agent.

REFERENCES

FRANK, J. D. *Persuasion and healing*. Baltimore: Johns Hopkins Press, 1961.
GORDON, P. Diphenylhydantoin and procaineamide: Normalization of suboptimal learning behavior. Paper presented at the meeting of the Society of Biological Psychiatry, Chicago, May 1967.
GORDON, P., TOSIN, S., DORV, B., & NASH, M. The effects of diphenylhydantoin and procaineamide on learning in animals and man. *Journal of Gerontology*, 1968, 23, 434-444.

- KATZ, M. M. Ethical issues in the use of human subjects in psychopharmacological research. *American Psychologist*, 1967, 22, 360-363.
- KUBIE, S. L. A psychoanalytic approach to the pharmacology of psychological processes. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960.
- KURLAND, S. A. Placebo effect. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960.
- PETERSON, E. Psychopharmacology. In J. A. Vernon (Ed.), *Introduction to psychology: A self selection textbook*. Dubuque, Ia.: Brown, 1966.
- RACHMAN, A. Affect-cognition. Unpublished doctoral dissertation, University of Chicago, 1965.
- ROGERS, C. R., & DYMOND, R. R. *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954.
- UHR, L. Objectively measured behavioral effects of psychoactive drugs. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960.
- WECHSLER, D. A standardized memory scale for clinical use. *Journal of Psychology*, 1945, 19, 87-95.
- WITTENBORN, J. R., & MAY, R. A. (Eds.) *Prediction of response to pharmacotherapy*. Springfield, Ill.: Thomas, 1966.
- ZIMRING, F. M. Some basic presuppositions of feeling and emotion. *Counseling Center Discussion Papers*, 1962, 8(7).
- ZIMRING, F. M. Towards a theory of feeling and emotion. *Counseling Center Discussion Papers*, 1963, 9(3).

(Received June 14, 1968)

TALKING TIME AS AN INDEX OF INTRAFAMILIAL SIMILARITY IN NORMAL AND ABNORMAL FAMILIES

WILLIAM D. WINTER¹ AND ANTONIO J. FERREIRA²

San Jose State College

In a recent study, Becker and McArdle found that members of normal families correlated more highly with each other in the average action or talking length of their extemporaneous speech than did members of clinic families. In the present research, 127 family triads (with varying diagnoses of the tested child) were asked to make up group Thematic Apperception Test stories, and the total number of seconds of speech of each family member was correlated with that of the other two family members. In contradiction to earlier findings, these results indicate that in this interactional situation members of abnormal families resemble each other more than do members of normal families. This is particularly true of the similarity between father and child, which was much greater in abnormal families.

In a recent study, Becker and McArdle (1967) asked individual family members to speak extemporaneously on some aspect of family life for 5 min. They found that normal family members resemble each other in the average duration of action length of their spontaneous speech more than do members of clinic families, presumably indicating greater identification in normals.

The authors had available data on the total number of seconds of speech, or talking time (TT), emitted by each of three family members who were *conjointly* making up three Thematic Apperception Test (TAT) stories (Ferreira, Winter, & Poindexter, 1966). The purpose of the present study was to determine whether normal and abnormal child families differ in the TT correlations among the triad tested.

METHOD

Subjects and Procedure

The families tested were 127 triads of father, mother, and child. These families were divided into four groups based on the diagnosis of the child: (a) 50 normals, (b) 44 emotionally maladjusted, (c) 16 schizophrenic, and (d) 17 delinquents. The children were at least 9½ yr. old, and the family members were white, literate, native-born Americans in good health, who had been living together as a family for more than 2 yr. Each family was pre-

sented with a set of three TAT cards and instructed to make up a story they could all agree upon, which would link the three cards together in a specified order, including in their story a description of what had led up to the situation, what the characters were doing and feeling, etc. (Winter, Ferreira, & Olson, 1965). This procedure took place three times, with three different sets of cards. The basic score used in the present study was TT emitted by each family member while all three stories were being made up by the family. The TT for each family member was intercorrelated with the TT for the other two members, and the correlations among the groups were compared.

RESULTS AND DISCUSSION

Table 1 shows the TT intercorrelations among the family triad for the normal and total abnormal groups as well as for various subsamples of the data, broken down by sex of child and specific abnormal diagnosis, both of which strongly affect the correlations. As can be seen from Table 1, for the sample taken as a whole the TT scores of the normal family members do not intercorrelate with each other, whereas the scores of all three dyads in abnormal families are significantly related. It would appear that for this non-lexical measure, at least, members of abnormal child families resemble each other, whereas members of normal families do not, a finding at variance with that of Becker and McArdle. This reversal of findings seems most evident in the father-child correlations, where the difference between the normal and abnormal groups is significant ($Z = 2.4$, $p < .02$). The

¹ Requests for reprints should be sent to William D. Winter, Counseling Center, San Jose State College, San Jose, California 95114.

² The authors express their appreciation to Ken Merritt for his help in analyzing these data.

TABLE 1

TOTAL TALKING TIME, IN SECONDS, CORRELATIONS
AMONG FAMILY MEMBERS, FATHER,
MOTHER, CHILD

Source	Father- mother	Father- child	Mother- child	N
Male + female				
Normals	.25	.09	.15	50
Abnormals	.38**	.49**	.26*	77
Emotionally malad- justed	.42**	.53**	.34*	44
Schizophrenic	.37	-.04	.18	16
Delinquent	.27	.56*	.15	17
Male child				
Normals	.07	.19	.17	34
Abnormals	.32*	.44**	.40**	55
Emotionally malad- justed	.38*	.37	.43*	28
Schizophrenic	-.19	.24	.48	11
Delinquent	.19	.60*	.17	16
Female child				
Normals	.64**	-.13	.06	16
Abnormals	.47*	.50*	.04	22
Emotionally malad- justed	.53*	.70**	.30	16
Schizophrenic	.52	-.53	-.10	5
Delinquent	—	—	—	1

* $p < .05$.

** $p < .01$, two-tailed.

authors used somewhat different measures of speech length with different samples engaged in different activities. However, it is important to realize that the conclusion of Becker and

McArdle that normal family members resemble each other in nonlexical speech more than do abnormal family members, thus implying greater identification with each other in normal families, may be limited to only certain kinds of situations, such as private, not interactional speech. From the present authors' data, equally plausible hypotheses could be formulated that in problem-solving interactional settings normal family members are free to talk or not talk as much as they please, while in abnormal families one or more members are more constrained to take their speech cues from another, or else they fall into a more rigid pattern of similar productivity based on habit, particularly for father and child. In any case, these hypotheses await further clarification.

REFERENCES

- BECKER, J., & McARDLE, J. Nonlexical similarities as an index of intrafamilial identifications. *Journal of Abnormal Psychology*, 1967, 72, 408-414.
- FERREIRA, A. J., WINTER, W. D., & POINDEXTER, E. Some interactional variables in normal and abnormal families. *Family Process*, 1966, 5, 60-75.
- WINTER, W. D., FERREIRA, A. J., & OLSON, J. L. Story sequence analysis of family TAT's. *Journal of Projective Techniques*, 1965, 29, 392-397.

(Received June 6, 1968)

ELECTROENCEPHALOGRAM AND RAPID EYE MOVEMENTS DURING FREE IMAGERY AND DREAM RECALL

JOSEPH REYHER¹ AND HOWARD MORISHIGE

Michigan State University

EOG, EEG, EMG, and heart rate of 10 male Ss were monitored under a visual imagery condition (free imagery) and a dream recall condition. Only the dream recall condition was associated with alpha desynchronization. The imagery of the dream recall condition was characterized by more primary process, more kinetic imagery, more frequent episodes of imagery, longer duration of images, and greater heart rate. Both conditions were associated with an equal increase in REMs, which were more characteristic of the waking state than those observed during nocturnal dreams. The differences between the conditions are discussed in terms of the nature of the tasks involved and the more blatant evidence of primary process regulation of the imagery in the dream recall condition.

Spontaneous visual imagery (free imagery) has been shown to have powerful uncovering properties in psychotherapy (Reyher, 1963, 1968). Resistances are acutely experienced by the client and primary process regulation of imagery progressively produces more blatant representation (derivatives and symbols) of repressed material. As the uncovering proceeds, anxiety and/or psychosomatic symptoms are produced along with even greater resistance. The client realizes from his own experience that something very disturbing inside of him is operating outside of his awareness. The goal of psychotherapy is the discovery of this pathogenic process. Eventually, if the client has sufficient ego strength, derivatives and visual symbols of repressed conflicts become directly depictive. No interpretations are required. This process is called "emergent" uncovering to distinguish it from the more traditional "interpretive" uncovering (Reyher, 1968, 1969).

Laboratory investigations into the psychodynamic properties of visual imagery are consistent with clinical experience. Burns (1967) found that the frequency of galvanic skin responses (GSR) was proportional to the amount of primary process intrusions in visual imagery. In another investigation Reyher and Smeltzer (1968) found that visual images in response to emotionally laden stimulus words

had greater uncovering properties than verbal associations. Visual imagery was associated with more frequent GSRs, more primary process, more direct representation of drives, and less effective defense than was verbal association.

The purpose of the present research was to monitor Ss while they were engaged in free imagery and dream recall to clarify the relationship between visual imagery, rapid eye movements (REMs), and the electroencephalogram (EEG). The authors were interested in whether the REMs of free imagery and recalled dreams resembled the REMs of waking imagery or the REMs of nocturnal dreaming. Dement and Kleitman (1957b) have hypothesized that REMs represent the dreamer's scanning of his visual images while Amadeo and Shagass (1963) view REMs as a nonspecific neural mechanism which is a concomitant of attentive activity. According to Shimazono, Ando, Sakamoto, Tanaka, Eguchi, and Nakamura (1965), REMs occurring during nocturnal dreams are generally greater in amplitude than waking REMs and take less time for deflection. Aserinsky and Kleitman (1955) noted that REMs during nocturnal dreams show relatively jerky motions with short arcs (generally .2 sec.) and tend to occur in clusters.

On the basis of the current literature, the authors did not expect the EEG to differentiate between conditions, but it was included as a way of providing information about Ss

¹ Requests for reprints should be sent to Joseph Reyher, Department of Psychology, Michigan State University, East Lansing, Michigan 48823.

state of awareness. Since the EEG findings turned out to be of considerable interest, salient aspects of the literature relating EEG characteristics to imagery are reviewed.

In 1934 Adrian and Matthews confirmed Berger's original findings that visual stimulation desynchronizes the alpha rhythm but that visual imagery does not. In contrast, Jasper and Cruickshank (1937) reported that the alpha rhythm is desynchronized during the formation of visual images in the flight of after images. They described a number of factors causing alpha desynchronization but concluded that the "attention value," "arousal value," or "reaction value" of the stimulus situation is the critical psychological variable. The alpha rhythm is controlled by an integrating subcortical pacemaker in the diencephalon which is a primary center of integration for nervous processes involving the autonomic nervous system.

Travis (1937) also found that the alpha rhythm was not desynchronized during visual stimulation, but he observed that the amplitude of the alpha waves were reduced for both visual images of light and objects, with the association appearing to be weaker for the latter. However, Golla, Hutton, and Walter (1943) distinguished among three groups of normal Ss in terms of alpha activity and reported that visual imagery was the most characteristic mental activity of those Ss with no alpha during eyes-closed rest. In 1954, Short and Walter contended that the desynchronization of the alpha rhythm in the waking state provides an objective index of the use of visual imagery in problem solving, but Barratt (1956) questioned this contention and showed that the amount of alpha desynchronization was unrelated to the degree of visual imagery that was involved. He concluded that alpha desynchronization is a function of attention.

Costello and McGregor (1957) improved upon Short and Walter's procedures and concluded that alpha desynchronization is the result of at least two factors: (a) the vividness of the visual image and (b) the extent to which thought associations, or what may collectively be called the "higher processes," are involved. They assumed that these higher

processes suppress visual imagery. In support of this contention, they referred to the blatant derivatives that are produced by a release in ego control, as noted by psychoanalysts and "Rorschach workers." They felt that the investigation of imagery and EEG changes is a worthy and fruitful endeavor in personality research.

More recent investigations do not remove the inconsistencies reported in the earlier literature. Foulkes and Vogel (1965) noted several stages of sleep onset that were characterized by vivid visual fantasies and alpha or theta waves; however, Stoyva and Kamiya (1968) reported that low alpha activity was associated with visual imagery in normal Ss and in Ss under the influence of LSD.

METHOD

Subjects

Eleven male college students ranging 18-29 yr. volunteered as Ss and all but one were readily able to experience visual images. The one exception was discontinued.

Scoring

REM. Voluntary eye movements in the waking state with eyes closed are easily detectable; however, involuntary eye movements are obscured by artifacts caused by muscle movements, brain waves, ocular tremors, and blinks. Eye tremors and blinks, for the most part, occur vertically and can be detected on the vertical electrooculograms (EOG). To reduce the effects of these artifacts, only eye movements monitored on the horizontal EOG were used to detect the occurrence of REMs.

Eye movements were recorded and measured by using procedures described by Dement (1955) and by Dement and Kleitman (1957a, 1957b). Conjugate horizontal eye movements were monitored by attaching a pair of Beckman biopotential electrodes at a point approximately 1 cm. lateral to the external canthus of each eye. Vertical eye movements and blinks were identifiable by means of supra- and infra-orbital biopotential electrodes on the left and right eyes, respectively. Bipolar potentials were recorded for both horizontal and vertical electrode pairs using an electrode attached to the mastoid area of the left ear as a ground.

According to Aserinsky and Kleitman (1955), the EOG contains slow (SEM) and rapid (REM) movements which are distinguished mainly in terms of duration and latency. Only REMs were considered in the present study, and difficulties in counting very small base-line oscillations and borderline SEMs were avoided by adopting the following criteria: (a) horizontal EOG deflection was equal to or greater in amplitude than the corresponding deflection on the

vertical EOG; (b) amplitude of the horizontal EOG was equal to or greater than 4 mm. (10 μ v.); and (c) duration from base to peak of .5 sec. or less. Waking REMs of 75 μ v. (30 mm.) (Caldwell & Domino, 1967a) or greater in amplitude and occurring in clusters were considered to be comparable to REMs occurring during nocturnal sleep.

EEG. The EEG was recorded bipolarly from the occipital area. Two silver cup electrodes filled with Bentonite electrode paste were attached to the skin by applying USP collodion. The positioning of the occipital EEG was done according to the International 10-20 Electrode System (Jasper, 1958). The S's earlobes wired in parallel served as a common reference electrode.

Due to the limitations involved in scoring attenuated alpha rhythm activity by visual inspection, occipital alpha was defined as the presence of not less than 3 consecutive waves of 8-13/sec frequency and 2 mm. (5 μ v.) or greater in amplitude. Percentage of time occipital alpha was defined as the duration of the total observed occipital alpha activity in seconds multiplied by 100 and divided by the length in seconds of the total duration of the episodes or period examined.

Heart rate. For interest, finger pulse was included to obtain an index of heart rate. The finger-pulse sensing device, provided by Ax and described in Ax, Beckett, Cohen, Frohman, Tourney, and Gottlieb (1962), was applied with elastic tape to the forefinger of S's nonpreferred hand.

The EEG, EOG, and finger pulse were monitored on the Grass No. 5 polygraph. Electrode resistances were checked prior to running S, and any electrode giving a reading above 10 K ohms was reapplied. The standard EEG constant speed of 10 mm/sec was used and the gain set so that a 20-mm. deflection was equivalent to 50 μ v.

Procedure

Training session. Each S was given a training session to provide him with experience in free imagery and to accustom him to the laboratory. He also was asked to abstain from all sedatives, stimulants, and alcoholic beverages at least 24 hr. prior to his experimental appointment which was scheduled for the following week.

Experimental session. The S was seated in a reclining chair in a dimly illuminated, soundproof room adjoining the equipment room where E was situated. The S was told that the purpose of the investigation was to determine the relationship between visual imagery, brain waves, and the electrical conductivity of the skin. No mention was made of eye movements. The E then attached the electrodes to S and asked him to close his eyes and keep as still as possible.

Free imagery condition. After an additional 5-min. period of adaptation to the laboratory situation, a trial run of 1-min. eyes-closed resting period (Base Line 1) was recorded to establish a base-line measure for later comparisons with visual imagery. The free

imagery condition began with a request for S to describe pictures that came into his mind's eye. He was asked to signal E (by pressing a switch with his preferred hand) the onset and termination of each episode of images. The switch, powered by a 3-v. dc battery, activated a signal light in the recording room. To avoid contaminating the imagery, S was asked to describe in detail what he saw after each imagery episode. The descriptions were recorded on a Uher 400 remote-controlled tape recorder with a built-in two-way intercom system. The free imagery session lasted for 45 min.

Owing to the long duration of the free imagery condition, there was some doubt concerning the meaningfulness of Base Line 1 throughout the entire period. Such processes as adaptation, fatigue, drowsiness, or increasing apprehension about the nature of the imagery could substantially change the percentage of alpha. Accordingly, the alpha percentage for each visual episode during the free imagery condition was compared with the alpha percentage in an equal period of no imagery just prior to the activation of the signal indicating the onset of visual imagery (Alternate Base Line 1).

Dream recall condition. On completion of the free imagery phase of the experiment, S was allowed to relax for 5 min. A second 1-min., eyes-closed resting period was monitored to establish a base line (Base Line 2) for this condition. After the trial run, S was instructed to close his eyes and visualize scenes from a single previous "vivid" nocturnal dream. As in the free imagery phase of the experiment, S was asked to signal the onset and termination of each visual imagery episode, and he was asked to describe what he saw after the entire dream was completed. The polygraph record of this phase lasted until the completion of S's "dream."

RESULTS

Visual Imagery

The total number of visual imagery episodes during the 45-min. free imagery condition of the experiment ranged 28-116, with an overall mean of 48.1, $SD = 27.6$, and a mean rate of 1.1 visual episodes per minute. The mean duration of each visual episode was 11.1 sec. In comparison, the mean duration of the dream recall condition was 1.4 min. and the total number of visual imagery episodes ranged 1-6, with an overall mean of 3.6, $SD = 1.7$, and a mean of 2.6 visual episodes per minute. The mean duration of each visual episode was 30.4 sec. The only question of significance in these data was the rate of visual episodes, and a t test showed that the two experimental conditions differed significantly (.05 level) in this regard. It also was obvious that the imagery of the dream recall

condition was more kinetic; that is, there was more movement in comparison to the free imagery condition wherein the images tended to be static. It must be emphasized, however, that there were some instances of highly kinetic imagery during the free imagery condition which is congruent with clinical experience with free imagery.

All of the Ss frequently showed a transient desynchronization of the alpha rhythm just prior (1 sec.) to signaling the onset of visual imagery. This transient initial desynchronization of alpha ranged between 46% and 100% with a mean of 60.1% for the free imagery condition, and between 1.7% and 100% with a mean of 58.2% for the dream recall condition.

As expected, most of the Ss did not experience primary process regulated imagery and/or the uncovering that often occurs with the clinical use of free imagery. For the most part, their imagery could be most aptly characterized as the visual reproduction of recent experience, everyday objects, and routine events of little obvious psychodynamic import, although it is known from clinical experience that such images often are really remote derivatives of unconscious strivings. This is apparent because over a session(s) images gradually become less symbolic and more representational, with the symbolic significance of previously "nonsymbolic" images becoming apparent. Regardless of the fact that this was an experiment with only 45 min. of free imagery experience, one S showed blatant primary process regulation of his imagery. In one episode he reported that he was peeking through the window of an old shop and an old man came in, pointed a finger at him and laughed, whereupon to S's consternation, the old man's finger turned into a penis. Of no little consequence was a concomitant increase in heart rate and a desynchronization of the alpha rhythm from 44% to 12%. If this S were a client in psychotherapy, he would have been asked to re-image the scene and the imagery would gradually become more depictive of a repressed striving. Anxiety and/or psychosomatic symptoms would have been experienced simultaneously with an increase in resistance

TABLE 1
MEANS OF REMs AND OCCIPITAL ALPHA FOR
EXPERIMENTAL AND BASE-LINE CONDITIONS

Episode	\bar{X}	SD	<i>t</i>
REM (mean REM/min)			
Base Line 1	1.0	.9	2.08*
Free imagery	2.9	3.1	
Base Line 2	1.6	3.1	1.99*
Recalled dream	3.3	5.4	
Free imagery	2.9	3.1	.33
Recalled dream	3.3	5.4	
Occipital alpha (mean % alpha)			
Base Line 1	33.0	23.2	.11
Free imagery	33.7	23.5	
Base Line 2	45.0	25.5	4.10**
Recalled dream	32.9	24.6	
Free imagery	33.7	23.5	.28
Recalled dream	32.9	24.6	

Note.—*N* = 10.

**p* ≤ .05, one-tailed.

***p* ≤ .005, two-tailed.

against the procedure. If the authors are correct, the old man eventually would have turned into S's father and at that moment he would have realized his sexual interest in his father's genitals.

REM

Table 1 shows the mean REMs/min under both experimental conditions, the results of which were analyzed separately by comparing each with its own base line. Both conditions were associated with significantly (.05 level) more REMs than their respective base lines; however, they did not differ from each other. Since the percentage of alpha of Base Line 1 and its alternate were almost identical (33.0 and 32.4), only the results of the former are reported.

REMs characteristic of nocturnal sleep were absent in both experimental conditions. The obtained REMs were less than the 75 μ v. criterion, were more gradual in slope, and did not occur in clusters. The REMs accompanying nocturnal dreams seem to have a component uniquely associated with the neural determinants of sleep, perhaps like those specified by Jouvet (1962). The recalled dreams seemed more like hypnogogic imagery than nocturnal dreams, according to the de-

TABLE 2
PERCENTAGE INCREASE AND DECREASE IN HEART
RATE FOR EXPERIMENTAL CONDITIONS

Episode	\bar{X}	SD	t
Within-experimental conditions			
Free imagery			
Decrease	19.2	12.0	
Increase	24.7	11.2	1.0
Recalled dream			
Decrease	11.6	17.6	
Increase	39.3	29.0	2.6*
Between-experimental conditions			
Decrease			
Free imagery	19.2	12.0	
Recalled dream	11.6	17.6	1.3
Increase			
Free imagery	24.7	11.2	
Recalled dream	39.3	29.0	1.6

Note.— $N = 10$.

* $p \leq .05$.

scriptions of Foulkes and Vogel (1965) and Roffwarg, Dement, Muzio, and Fisher (1962). The reimagining of a nocturnal dream had a mean duration of 2 min. in contrast to means of 9 and 11.2 min. reported by Dement and Kleitman (1957a) and Caldwell and Domino (1967b), respectively. There was also an absence of visual continuity of images during the recalled dream.

EEG

The results of Table 1 indicate that alpha activity during free imagery did not decrease; however, there was alpha desynchronization during the visual imagery episodes in the dream recall condition relative to Base Line 2 (.005 level). This result is due to a gross difference between the two base lines, not because there was a difference between the experimental conditions.

Heart Rate

The onset of each visual episode was examined for changes in heart rate. Table 2 shows that there was a significant increase in heart rate for the dream recall condition but not for the free imagery condition. Although the remaining comparisons were uniformly non-significant, they were in the right direction and further suggest that the dream recall condition had more central nervous system (CNS) activating properties or, in psycho-

dynamic terms, more anxiety-producing properties.

DISCUSSION

The desynchronization of the alpha rhythm during the dream recall condition, relative to its base line, was surprising in view of the lack of alpha desynchronization for the free imagery condition, relative to its base line. Inspection of Table 1 reveals that the relative desynchronization of the dream recall condition is solely due to the increase of alpha for Base Line 2. Since the percentage of alpha for the two experimental conditions was the same, it might be argued that there is an upper limit of alpha activity during visual imagery. This hypothesis is vitiated, however, by the wide range of percentage of alpha (2.8 to 74.6) in the free imagery condition, which was comparable to that of Base Line 1 and its alternate. Another hypothesis is that the percentage of alpha in the free imagery condition was too low to be desynchronized by visual imagery; however, the obtained percentage of alpha in this condition is consistent with the percentage of alpha in an eyes-closed resting period often cited in the general EEG literature. Moreover, according to Ostow (1950), the direction of change is unrelated to the base-line percentage of alpha. There should have been some desynchronization regardless, but there was not.

The so-called law of initial values (LIV), which specifies a positive correlation between the initial level of a function and the amount of change from that level, is pertinent to Base Line 2 and the dream recall condition. For these data there should be relatively greater desynchronization of alpha for Ss with high initial alpha (Base Line 2) resulting in a sharply reduced mean percentage of alpha for the dream recall condition. An inspection of the scores of individual Ss between Base Line 2 and the dream recall condition reveals that only one S did not show a decrease in alpha, and the correlation (Spearman) between initial level and amount of change was only .22. The LIV does not appear to be an influencing variable.

There are two viable hypotheses yet to be considered. The first would attribute the relative desynchronization of the dream recall

condition to the differential demands made by the two experimental conditions. In the free imagery condition, *S* is not asked to do anything other than press a button at the onset of a visual image, whereas in the dream recall condition he is asked to do something specific. He must make the effort to recall a dream. It is this effort which creates a pattern of CNS activation of which reduced alpha is one aspect. The second hypothesis takes into account the activating properties of primary process thinking which was much greater for the dream recall condition. It was the anxiety implicit in primary process thinking that caused CNS activation and alpha desynchronization. Related investigations in the authors' laboratory using the GSR (Burns, 1967; Reyher & Smeltzer, 1968) show that primary process regulated visual imagery is related to CNS activation of the GSR. Other studies, yet incomplete, are revealing the same relationship, and further research is being planned to separate and evaluate these two variables.

Taken together, these hypotheses (task and primary process sources of CNS activation) might account for the contradictory findings in the research literature concerning the relationship between alpha and visual imagery. Unlike the present investigation, previous investigations did not monitor all *S*'s visual imagery; either *S* was asked to describe his mental state because of some change in the EEG and/or in REMs, or he was asked to visualize something in particular. In the first case, only visual imagery which had activating properties would have been detected, and in the latter case, it is the request per se to form a specific visual image which itself is activating, not the image. Because of these biasing circumstances, it is understandable why the research literature is confusing.

The data also have implications with respect to the hypothesis that REMs are related to *S*'s scanning of his images. If this were true, there should have been more REMs during visual imagery in each experimental condition relative to its base line, which was the case. Also, there should have been an increase in REMs for the experimental condition associated with more of any one of the following characteristics: (a) higher rate of

visual episodes, (b) longer duration of visual episodes, and (c) kinetic imagery. Since the dream recall condition had a greater association with all three of these characteristics in comparison with the free imagery condition, it should also be associated with more REMs, which was not the case. These results are more consistent with the hypothesis advanced by Amadeo and Shagass (1963); that is, REMs represent a nonspecific neural mechanism which is a concomitant of attentive activity. Accordingly, *Ss* in the present research paid equal attention to the two experimental conditions, even though they differ in terms of CNS activation as suggested by the data on heart rate. Is there perhaps an upper limit to attention? Attention is a state of awareness in which the individual is tuned to respond adaptively to sensory input. Attention can be reduced below this level by fatigue, drugs, or loss of interest in the sensory stimuli, but it cannot be raised above this level.

The data of the *S* who saw the old man in the window are pertinent to the hypothesis advanced by Antrobus and Singer (1965) which attributes an increase in REMs to the suppression of fantasies. He saw a snake 5.5 sec. after seeing the old man's finger turn into a penis, indicating that anxiety associated with the old man and his penis finger was excessive and that repressive defenses were stimulated to produce a less blatant derivative. This heightened level of defense is maintained over the following images: flag pole with soldiers marching, a huge missile going by, and a girl's face with no features; but no increase in REMs was noted. Until more data like this are collected, which is common in emergent uncovering, it is concluded very tentatively that REMs are not associated with an increase in defense. Antrobus and Singer asked their *Ss* to suppress, and it may be the effort involved in this task that is associated with REMs, not the presence or absence of visual imagery.

Free imagery has not only proved itself to be a powerful uncovering technique in psychotherapy, but it is showing promise as a research tool as well. Its value as a research tool inheres in the response-producing psychodynamic processes that are set in motion and

rendered visible in the form of derivatives and symbols of repressed drives and conflicts which gradually become less symbolic and more depictive. The objectivity of the data can be preserved if interpretations are not given to the client.

REFERENCES

- ADRIAN, E. D., & MATTHEWS, B. The Berger rhythm: Potential changes from the occipital lobes in man. *Brain*, 1934, 57, 355-385.
- AMADEO, M., & SHAGASS, C. Eye movements, attention, and hypnosis. *Journal of Nervous and Mental Diseases*, 1963, 136, 139-145.
- ANTROBUS, J., & SINGER, J. Eye movements during fantasies: Imagining and suppressing fantasies. *Archives of General Psychiatry*, 1965, 12, 71-76.
- ASERINSKY, E., & KLEITMAN, N. Two types of ocular motility occurring in sleep. *Journal of Applied Physiology*, 1955, 8, 1-10.
- AX, A., BECKETT, P., COHEN, B., FROHMAN, C., TOURNEY, G., & GOTTLIEB, J. Psychophysiological patterns in chronic schizophrenia. *Recent Advances in Biological Psychiatry*, 1962, 4, 218-233.
- BARRATT, P. Use of the EEG in the study of imagery. *British Journal of Psychology*, 1956, 47, 101-114.
- BURNS, B. The effect of repression on visual imagery. Unpublished masters thesis, Michigan State University, 1967.
- CALDWELL, D., & DOMINO, E. Electroencephalographic and eye movement patterns during sleep in chronic schizophrenic patients. *Electroencephalography and Clinical Neurophysiology*, 1967, 22, 414-420. (a)
- CALDWELL, D., & DOMINO, E. Patterns of sleep in schizophrenia. In G. Tournay & J. Gottlieb (Eds.), *The Lafayette clinic studies on schizophrenia*. Detroit, Mich.: Wayne State University Press, 1967. (b)
- COSTELLO, C., & MCGREGOR, P. The relationships between some aspects of visual imagery and the alpha rhythm. *Journal of Mental Science*, 1957, 103, 786-795.
- DEMENT, W. Dream recall and eye movements during sleep in schizophrenics and normals. *Journal of Nervous and Mental Diseases*, 1955, 122, 263-269.
- DEMENT, W., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 1957, 9, 673-690. (a)
- DEMENT, W., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method of the study of dreaming. *Journal of Experimental Psychology*, 1957, 53, 339-346. (b)
- FOULKES, D., & VOGEL, G. Mental activity at sleep onset. *Journal of Abnormal Psychology*, 1965, 70, 231-243.
- GOLLA, F., HUTTON, E. L., & WALTER, W. G. The objective study of mental imagery. I. Physiological concomitants. *Journal of Mental Science*, 1943, 89, 216-222.
- JASPER, H. Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, 1958, 10, 370-375.
- JASPER, H. H., & CRUIKSHANK, R. M. Electroencephalography: II. Visual stimulation and the after-image as affecting the occipital alpha rhythm. *Journal of General Psychology*, 1937, 17, 29-48.
- JOUVET, M. Recherches sur les structures nerveuses et les mecanismes responsables des differentes phases de sommeil physiologique (Investigations of the nervous structures and the mechanisms responsible for different phases of physiological sleep). *Archives of Italian Biology*, 1962, 10, 125-206.
- OSTROW, M. Psychic function and the electroencephalogram. *Archives of Neurology and Psychiatry, Chicago*, 1950, 64, 385-400.
- REYHER, J. Free imagery: An uncovering procedure. *Journal of Clinical Psychology*, 1963, 19, 454-459.
- REYHER, J. *Hypnosis*. Dubuque, Ia.: Brown, 1968.
- REYHER, J. Emergent uncovering in psychotherapy and the objectification of psychodynamic processes. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September 1969.
- REYHER, J., & SMELTZER, W. The uncovering properties of visual imagery and verbal association: A comparative study. *Journal of Abnormal Psychology*, 1968, 73, 218-222.
- ROFFWARG, H., DEMENT, W., MUZIO, J., & FISHER, C. Dream imagery: Relationship to rapid eye movements of sleep. *Archives of General Psychiatry*, 1962, 7, 235-258.
- SHIMAZONO, Y., ANDO, K., SAKAMOTO, S., TANAKA, T., EGUCHI, T., & NAKAMURA, H. Eye movements of waking subjects with closed eyes. *Archives of General Psychiatry*, 1965, 13, 537-543.
- SHORT, P. L., & WALTER, W. G. The relationship between physiological variables and stereognosis. *Electroencephalography and Clinical Neurophysiology*, 1954, 6, 29-44.
- STOFFA, J., & KAMIYA, J. Electrophysiological studies of dreaming as the prototype of a new strategy in the study of consciousness. *Psychological Review*, 1968, 75, 192-205.
- TRAVER, L. E. Brain potentials and the temporal course of consciousness. *Journal of Experimental Psychology*, 1937, 21, 302-309.

(Received July 8, 1968)

EFFECTS OF CONDITIONING UPON PREEXPERIMENTALLY ACQUIRED VERBAL HABIT IN PSYCHIATRIC PATIENTS

MICHAEL L. HURT AND CLAIRE FALK ETAUGH¹

Bradley University

The Ss were 20 adult females, hospitalized for psychiatric disorders, with 7 displaying initial negative verbal habit (i.e., using bad-impression verbs with self-referent pronouns) and 13 displaying initial positive verbal habit (i.e., using good-impression verbs with self-referent pronouns). The Ss were matched for diagnostic category, age, education, and number of hospitalizations, and divided into two groups. Experimental Ss reinforced for displaying a verbal habit opposite to their initial habit significantly increased their tendency to display the reinforced habit. When the initial verbal habit was then reinforced, Ss significantly increased their tendency to display this habit. Control Ss never received reinforcement, and showed no change in initial verbal habit.

Some recent research has focused on the influence of previously learned verbal habits on verbal conditioning. Dixon (1965) studied the role of the impression value of verbs in the conditioning of the pronouns "I" and "we." He employed a version of Taffel's (1955) sentence-construction task, which requires S to choose one of six pronouns as the first word of a sentence whose second word is a past-tense verb. Dixon found that even in the operant period the pronouns I and we were used more frequently with good-impression verbs than with bad-impression verbs, indicating that such verbal habits are acquired prior to experimental reinforcement. Moreover, reinforcement of I and we increased not only the frequency of these pronouns, but also the use of good-impression verbs with these pronouns. Clance and Dixon (1965) found that the influence of verbs on pronoun choice extends to other-referent, as well as self-referent, pronouns. Both studies indicated that prior associations with verbs are greater determinants of pronoun choice than is currently administered reinforcement.

The present study concerns the extent to which such previously acquired verbal habits can be modified. The distinction should be made between positive verbal habit (PVH) and negative verbal habit (NVH). The former may be defined as the tendency to associate good-impression verbs with self-referent pro-

nouns; the latter, as the tendency to associate bad-impression verbs with such pronouns. Since it is possible that PVH and NVH are differentially affected by reinforcement operations, it was felt desirable to study both habits. Normal Ss, however, habitually display only PVH (Clance & Dixon, 1965; Dixon, 1965). The Ss for the current study were selected from a population of psychiatric patients since, unlike the normal populations previously studied, it included individuals who exhibited NVH, as well as those who showed PVH (unpublished pilot study by the senior author).

The hypothesis tested was that Ss reinforced for displaying a verbal habit (either positive or negative) opposite to their initial verbal habit (either positive or negative) increase their tendency to display the reinforced habit. Further, when the reinforcement contingency is shifted so that the original verbal habit is reinforced, Ss should increase their tendency to display the original habit.

METHOD

Subjects. Twenty females hospitalized for psychiatric disorders were matched on diagnostic category, age, education, and number of hospitalizations, and divided into experimental and control groups. The mean chronological age of both the experimental and control groups was 40.2 yr. The mean number of years of formal education for the experimental and control groups was 10.7 and 10.4, respectively. The mean number of previous hospitalizations for the experimental group was 5.1, and for the controls, 5.5. On the operant trial of the experi-

¹ Requests for reprints should be sent to Claire Falk Etaugh, Department of Psychology, Bradley University, Peoria, Illinois 61606.

mental task, five experimental Ss displayed PVH and five displayed NVH; for the controls, these figures were eight and two, respectively.

Task. In a modified version of the Taffel (1955) task, the pronoun I was presented with a selection of 30 past-tense verbs that had been scaled previously for impression value (Dixon & Dixon, 1964) and employed by Dixon (1965). Ten verbs connoted a good impression; 10, a neutral impression; and 10, a bad impression. Each of the 150 index cards contained the pronoun I and 2 verbs centered beneath it. Verbs in each of the three impression-value classes appeared two-thirds of the time, and were paired equally often with verbs from the other two categories. The 30 verbs were so paired that the same 2 verbs did not appear together more than once.

Procedure. After being told that she was to participate in a personality study, S was instructed to construct a sentence for each card, using I as the first word and one of the two verbs as the second. Each set of 15 cards constituted a trial, with the first trial establishing the base level of response. During Trial 1, E made no response to S. During Trials 2 through 6 with the experimental group, E reinforced those responses opposite to the base level response tendency; that is, those Ss displaying PVH during the operant trial were reinforced for displaying NVH, and vice versa. During Trials 7 through 10, E reversed the reinforcement contingency and supported the original verbal habit of the experimental Ss. The E reinforced desired responses by smiling, nodding his head, and uttering one of the following statements: "You are doing a good job"; "That's fine"; "That was a good sentence"; "Very good." Control Ss received no reinforcement during the 10 trials.

Following the experiment, each S answered an awareness questionnaire to determine if she could state the reinforcement contingencies (see Dixon, 1965).

RESULTS

The principal results are shown in Figure 1, which presents the mean numbers of initial and opposite verbal habit responses as a function of trials for experimental and control Ss.

McNemar's (1955) test for significance of change was used to evaluate the data. In line with the hypothesis, the experimental Ss significantly changed from their initial verbal habit to the opposite habit when the latter was reinforced ($\chi^2 = 8.10$, $p < .01$), and then showed a statistically significant return to their initial verbal habit when the reinforcement contingency was reversed ($\chi^2 = 7.01$, $p < .01$).

The data for the experimental group were analyzed separately for PVH and NVH Ss by means of the binomial test (Siegel, 1956). All

Ss in both subgroups displayed the predicted change to the opposite verbal habit when it was reinforced ($p < .03$, in both cases). During the reversal phase, all the original PVH Ss returned to their initial verbal habit ($p < .03$), and all but one of the original NVH Ss did so. This latter finding is statistically non-significant, given the smallness of the sample size ($n = 5$).

Control Ss showed no significant change in verbal habit throughout the 10 trials.

None of the Ss indicated knowledge of the reinforcement contingencies on the awareness questionnaire.

DISCUSSION

The results indicate that in a highly structured situation, the habitual pattern of associating a specific class of verbal impression values with self-referent pronouns may be changed and even reversed through the application of reinforcement. This finding holds true irrespective of the nature of the original verbal habit (i.e., positive or negative). These results are relevant to Ellis' (1958) hypothesis that changes in verbal behavior are essential in producing changes in emotional reactions. He contends that one may substantially change one's feelings by changing the internalized verbalized behavior by which one initially created the feelings. If, as Ellis believes, emotional disturbance essentially arises when individuals repeat to themselves negative, self-defeating statements, then the substitution of more positive self-referent statements should be correlated with a cessation of the disturbance.

The efficacy of reinforcement in modifying verbal behavior of psychiatric patients in this study stands in contrast to the findings of Dixon (1965) and Clance and Dixon (1965), in which prior verbal habits appeared to be greater determinants of verbal response than were current reinforcement contingencies. The apparent contradiction may be accounted for in terms of two variables: (a) the nature of the reinforcers and (b) S characteristics. All three studies used social reinforcement. However, the Dixon studies employed only the spoken word "good," whereas the present study used one of a sequence of verbal reinforcers, accompanied by a smile and head nod.

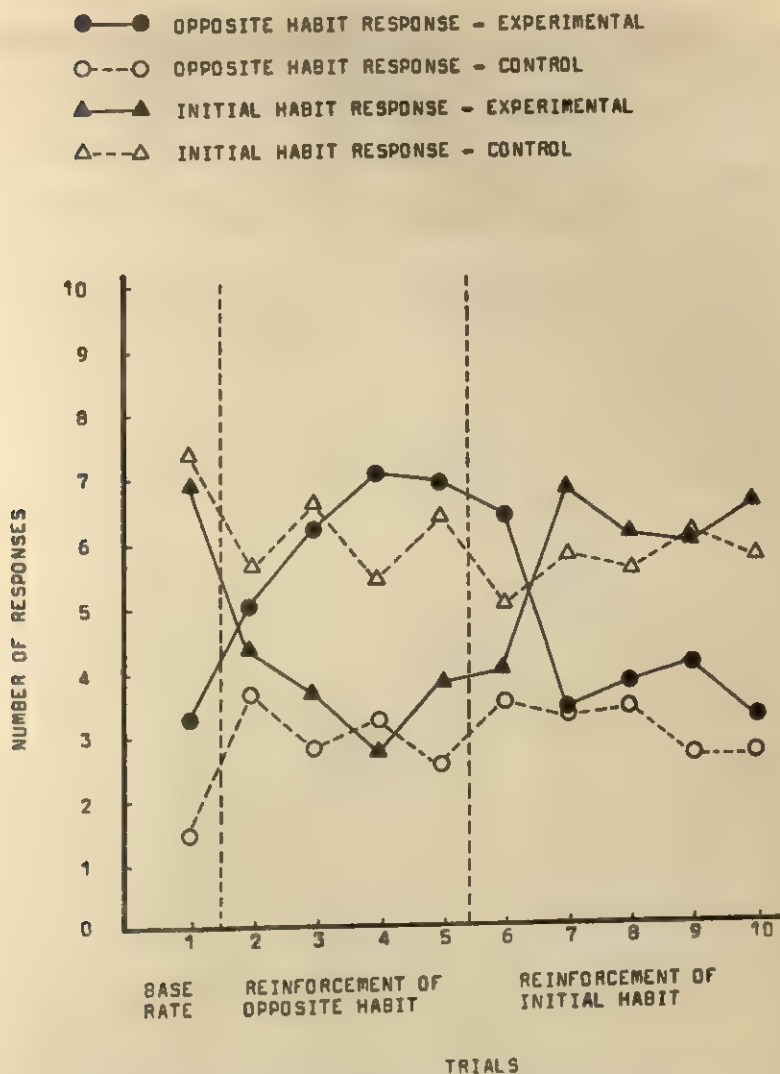


FIG. 1. Mean numbers of initial and opposite verbal habit responses as a function of trials for experimental and control Ss (combined data for PVH and NVH Ss).

Wickes (1956), comparing a sequence of verbal reinforcers with a sequence of gestural reinforcers, found that both increased the frequency of *M* responses on the Rorschach test, with the nonverbal reinforcement being somewhat more effective. It is likely that the combination of verbal and gestural reinforcers used in the present study were more effective than the single word good used by Dixon. In addition, Ss might be less likely to satiate on

a variety of reinforcers than on one consistently presented.

With regard to differences in *S* characteristics, the Ss in the present study were institutionalized, whereas Dixon's were not. The finding that social reinforcement was apparently more effective with the former Ss is in line with reports of Zigler (1963) and others that institutionalization enhances the effectiveness of social reinforcement.

REFERENCES

- CLANCE, P., & DIXON, T. P. The effect of previously learned verbal habits upon two response classes in verbal conditioning. *Journal of Psychology*, 1965, 60, 271-276.
- DIXON, T. The effect of pre-experimentally acquired verbal habits upon verbal conditioning. *Journal of Psychology*, 1965, 59, 335-347.
- DIXON, T., & DIXON, J. F. The impression value of verbs. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 161-165.
- ELLIS, A. Rational psychotherapy. *Journal of General Psychology*, 1958, 59, 35-49.
- McNEMAR, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.
- SIEGEL, S. *Nonparametric statistics: For the behavioral sciences*. New York: McGraw-Hill, 1956.
- TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, 51, 496-500.
- WICKES, T. A. Examiner influence in a test situation. *Journal of Consulting Psychology*, 1956, 20, 23-25.
- ZIGLER, E. F. Rigidity and social reinforcement effects in the performance of institutionalized and noninstitutionalized normal and retarded children. *Journal of Personality*, 1963, 31, 258-270.

(Received July 10, 1968)

SYSTEMATIC DESENSITIZATION VERSUS IMPLOSIVE THERAPY¹

CURTIS L. BARRETT²

Child Psychiatry Research Center

University of Louisville School of Medicine

Systematic desensitization (SDT) and implosive therapies (IT) were compared for their effectiveness and efficiency in reducing snake phobic behavior in otherwise normal adult human beings. SDT and IT Ss differed significantly from control Ss in posttreatment avoidance of a snake and in change of reported discomfort. SDT and IT, however, did not differ in effectiveness. These results held at a 6-mo. follow-up. IT was more efficient in that treatment was completed in 45% of the time required for SDT. Results were qualified by the finding that SDT had a consistent and continuing effect across Ss and across time whereas IT effect was more variable. Occurrence of unusual disturbance of Ss during SDT and of IT Ss between sessions was discussed.

Two current learning-theory-based therapies, applied to reduction of phobic behaviors, call for directly opposed operations. These are Wolpe's (1958) systematic desensitization therapy (SDT) and Stampfl's (1961) implosive therapy (IT). In Wolpe's well-known approach, relaxation is used in reciprocal inhibition of small doses of anxiety that may be aroused by a hierarchical arrangement of stimuli. Since relaxation is assumed to inhibit only small amounts of anxiety, a gradual progression through the hierarchy is used. Considerable care is taken to see that Ss experience little or no anxiety as they imagine elements of feared stimulus events.

Stampfl's approach is the opposite of Wolpe's. He assumes that extinction of a negative emotional response to objectively neutral stimuli will proceed most rapidly when

Ss are exposed to a conditioned stimulus (CS) without reinforcement. Use of a gradual approach, for Stampfl, invites reinstatement of the response to the CS in accordance with the conservation of anxiety hypothesis (Solomon, Kamin, & Wynne, 1953). Therefore, as Stampfl and Lewis stated, IT therapists attempt to "represent, reinstate, or symbolically reproduce the stimuli (cues) to which the anxiety response has been conditioned, in the absence of primary reinforcement [p. 499]." This is accomplished by instructing S to imagine scenes suggested by the therapist until some sign of reduction in the anxiety elicited by the scene occurs. Then scenes that produce even more anxiety are suggested. This process is continued until the scenes suggested by the therapist no longer elicit anxiety.

A practitioner's choice of either SDT or IT on theoretical grounds would be premature since both therapies must await future demonstration of the adequacy of their underlying theories. Choice on the basis of effectiveness also would be difficult since both SDT and IT, when studied under well-controlled conditions, have been shown to be effective. Lang, Lazovik, and Reynolds (1965), for example, found SDT to be more effective than control or "pseudotherapy" procedures in reducing snake phobic behavior. Paul (1966) showed SDT to be more effective than insight therapy or control-placebo procedures in reducing performance anxiety. Similarly, IT has been

¹ This paper is based on a dissertation submitted to the Department of Psychology, University of Louisville in partial fulfillment of the requirements for the PhD degree. The research was supported in part by California Department of Mental Hygiene, Patton State Hospital Research Project PSH-141. The author wishes to express sincere appreciation to Roger M. Gardner, who served as *E_s*, to Glynn D. Coates, Ronald S. Lindner, and Lovick C. Miller, who served as therapy session raters, to Jack Fox and Walter W. Surwillo for their guidance and valued comments, and to Willard A. Mainord, dissertation advisor, who also served as the clinical psychologist for the study.

² Requests for reprints should be sent to the author, Child Psychiatry Research Center, 608 South Jackson Street, Louisville, Kentucky 40202.

shown to reduce rat phobic behavior more than control procedures (Kirchner & Hogan, 1967) and to be effective in the treatment of hospitalized schizophrenics (Hogan, 1966) and neurotic outpatients (Levis & Carrera, 1967). But there has been no direct comparison of SDT and IT such as that conducted by Paul (1966) to compare SDT and insight therapy.

The present study was conducted (a) to provide a direct comparison of the effectiveness of SDT and IT and (b) to compare the efficiency of SDT and IT in reducing the same disordered behavior. Fear of harmless snakes, as it occurs in otherwise normal adult human beings is the disordered behavior subjected to study. Efficiency of a therapy, for purposes of this study, is defined as time required for treatment. Thus, of two equally effective therapies, the one that takes less time is regarded as more efficient.

METHOD

Thirty-six Ss, 8 males and 28 females, were carefully selected according to the following criteria: (a) demonstrated fear of a harmless snake, (b) absence of severe emotional disorder, (c) verbalized reason to overcome fear of harmless snakes, and (d) commitment to participate in up to 15 hr. of therapy.

Fear Measures

Several measures of fear were administered by E_s , who participated in no other aspect of the experiment.

Avoidance test (AT). The a priori criterion measure for fear reduction, and therefore, for therapy effectiveness, was S 's passing of an AT used by Lang et al. (1965). Passing required that S touch or hold a 6-ft. black snake.

Avoidance test change score (ATCS). The AT yielded a percentage change statistic (ATCS) (Lang et al., 1965) based on S 's approach behavior. Scores ranged from 1 for holding the snake; 2, touching it; 3, standing next to the cage; 18, standing 15 ft. away; to 19, refusing to enter the room. The ATCS equaled S 's pre-AT score minus his post-AT score divided by his pre-AT score.

Fear survey schedule (FSS III). Wolpe and Lang's (1964) fear survey schedule was used to assess S 's reported reaction to a variety of objects and situations using a 5-point scale (not at all, 1; to very much, 5). The sum of S 's responses (FSS III, total score) indicated his generalized disturbance. Item 63 of the FSS III refers to harmless snakes.

Fear thermometer (FT). Walk's (1956) fear thermometer permitted S to rate degree of fear felt in

the test situation using a thermometer-shaped device with 10 divisions.

S-R inventory of anxiousness (SRIA). The Endler, Hunt, and Rosenstein (1962) SRIA permitted S to respond to 14 items describing his response to being asked to pick up a live, harmless snake. The sum of S 's responses on a 5-point scale, to 14 items, served as an index of anxiety that S experienced when confronted with the snake.

Procedure

Persons who were afraid of harmless snakes were sought in the Louisville area from graduate and undergraduate classes, nursing schools, and from the public via newspaper publicity. Those whose fear appeared to be genuine received pretreatment testing (AT, FSS III, FT, and SRIA) which was conducted as follows. E_1 introduced S to E_2 who asked S to come to a room in which, 15 ft. from the door, there was a caged, live 6-ft. black snake. The S was asked successively to approach, to touch, and to hold the snake. Afterwards, S completed the other fear measures and was interviewed by E_2 in order to tape-record S 's qualitative statements of the experience. The S s who failed AT, that is, those who neither held nor touched the snake, were next given the Minnesota Multiphasic Personality Inventory (MMPI). Those whose MMPI profile appeared to be normal were referred to an experienced clinical psychologist (CP). If CP found no evidence of present or past emotional disorder other than snake phobia, S was accepted and assigned randomly to one of three groups: systematic desensitization therapy, implosive therapy, or control group.

Control group (CG). The S s assigned to CG were told that they would be contacted concerning the next step in their "therapy." Next, CG S s received a letter that arranged for posttreatment testing after 8 wk. The CG S s were contacted similarly to arrange for a follow-up testing after 6 mo. After follow-up testing, E_2 explained the experimental procedures and the function of control groups. Then he offered S s treatment for the phobia at the University of Louisville Psychology Clinic.

SDT. The S s assigned to SDT received 4 training sessions and up to 11 SDT sessions. In the first training session, S s heard a tape recording which briefly explained SDT theory and which indicated that the investigators, as scientists, would be pleased no matter what the results of the experiment. The remainder of Session 1 and the next 3 sessions were used to construct an individual hierarchy for S by Lang and Lazovik's method (1963) and to teach S s to relax. Relaxation instructions were based on Paul (1966, pp. 118-120). The SDT proper was begun in Session 5 and essentially was conducted as described by Paul (1966, pp. 120-122). The criterion for termination of SDT in fewer than 11 treatment sessions was presentation of the hierarchy twice without S signalling disturbance. The SDT S s received the post- and follow-up testing that CG S s received.

TABLE 1
SUMMARY OF RESULTS AT POSTTREATMENT TESTING

Group	AT passed* (<i>N</i> = 12)	ATCS		FT change		FSS III				SRIA total score change	
						Item 63 change		Total score change			
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	1	.13	.35	.92	1.78	.66	2.15	4.16	19.63	2.92	5.32
SDT	11	.92	.17	4.25	2.47	2.33	1.31	31.67	29.32	12.33	9.57
IT	10	.88	.30	4.33	2.43	2.42	1.24	28.33	16.23	15.58	10.48

Note.—SDT = systematic desensitization therapy; IT = implosive therapy; AT = avoidance test; ATCS = avoidance test change score; FT = fear thermometer; FSS III = fear survey schedule; and SRIA = S-R inventory of anxiousness.

* Passing the avoidance test means *S* either touched or held the snake.

IT. The *Ss* assigned to IT received two 50-min. clinical interviews followed by up to two sessions of IT which were of unspecified duration. In the first interview, IT *Ss* heard a recorded explanation of IT that was identical to the SDT rationale tape except for the portion describing IT. Interviews I and II were also used to explore *Ss*' fear of snakes. Near the end of the second interview *S* practiced neutral imagery such as walking into a store. Interviews III and IV were conducted as described elsewhere by Stampfl (1961), Stampfl and Levis (1967), and Barrett (1967). Essentially, a scene such as confronting a snake while climbing a tree, which initially was thought to arouse maximum anxiety in *S*, was presented until *S* showed some diminution of anxiety. Then a new scene (e.g., being enveloped in a python's coils) was presented until it too ceased to elicit anxiety. Scenes were tailored to the experience of each *S* but scenes depicting helplessness, bodily harm, envelopment, and being devoured by huge snakes were commonly used. The only departure from standard IT procedure was exclusion of psychodynamic material. The termination criterion for IT was the same as that for SDT: presenting *Ss*' most frightening scene twice without his indicating disturbance ("fear," "anxiety"). The IT *Ss* received the post- and follow-up testing received by SDT and CG *Ss*.

Therapy ratings. All SDT and IT sessions were tape-recorded and then rated by two clinical psychologists and one experimental psychologist (all PhDs, one ABPP) who were trained in SDT and IT for the project. The rating procedure required identification of randomly selected 5-min. excerpts as SDT or IT and indication of degree of adherence to SDT or IT operations. A rating of 1 indicated no adherence to operations (therapy unacceptable); 7 indicated that operations were followed exactly. Acceptance of an excerpt by two of three raters admitted the results of the therapy to the study. To offset any bias of the excerpts, raters were provided with three randomly selected complete interviews for each SDT and IT *S* (72 recordings). Raters were asked to sample these in order to satisfy themselves that the excerpts were representative of actual therapies.

RESULTS

Results of this study necessarily hinge on the demonstration that *Ss* received the therapy that was appropriate for their group. Raters were able to identify all SDT excerpts as SDT and all IT excerpts as IT. Additionally, the mean rating for SDT sessions ($\bar{X} = 6.33$, $SD = .66$) and IT sessions ($\bar{X} = 6.10$, $SD = .91$) did not differ ($t = 1.15$, $df = 70$, $.10 < p < .20$). All three raters agreed on a rating of 5 or higher in 91.6% of SDT cases and in 83.3% of IT cases. Thus, there was evidence to indicate that rater agreement was quite high, the therapies were correctly identified by raters, and the degree of adherence to operations did not differ for SDT and IT.

Posttreatment findings are summarized in Table 1. Fisher's exact probability test indicated that the frequency of passes of AT for SDT and IT differed from that for CG ($p < .01$), but SDT and IT did not differ. Analysis of variance followed by orthogonal comparison (Winer, 1962, p. 66) showed similar results for ATCS and for simple change scores (pre- minus post-) that were computed for the remaining measures. That is, SDT and IT differed from CG ($p < .01$) but not from each other.

Table 2 shows the results for 6-mo. follow-up. The same procedures of analysis that were used for postresults, performed on frequency of passes on the AT and change on FT; on FSS III, total scores; on SRIA, Item 63; and on SRIA, showed that SDT and IT differed from CG ($p < .05$). However, SDT and IT did not differ in fear reduction

TABLE 2
SUMMARY OF RESULTS AT 6-MONTH FOLLOW-UP

Group	AT passed ^a (<i>N</i> = 12)	ATCS		FT change		FSS III				SRIA total score change	
						Item 63 change		Total score change			
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	3	.22	.57	1.79	1.96	.25	.87	5.16	23.59	6.33	6.89
SDT	12	.97	.06	3.75	3.34	2.25	1.29	33.42	27.17	14.00	8.48
IT	9	.79	.39	5.00	1.99	2.83	1.27	35.42	29.66	17.58	9.61

Note.—See Table 1 for explanations of abbreviations.

^a Passing the avoidance test means *S* either touched or held the snake.

assessed by these measures. Follow-up ATCS data were assessed by *t* tests since the hypothesis of homogeneity of variance for these data was rejected (Cochran's *C* > .61). Follow-up ATCS for CG differed from that for SDT ($t = 4.68$, $df = 22$, $p < .01$) and IT ($t = 2.85$, $df = 22$, $p < .01$) but SDT and IT did not differ ($t = 1.63$, $df = 22$, $.10 < p < .20$).

Table 3 shows mean training and treatment time for SDT and IT. Considering treatment time only, IT was completed in 41% of the time required for SDT. Comparing treatment plus training time, IT was completed in 45% of the time required for SDT.

The original experimental design of this study, as shown in the preceding paragraphs, called for evaluation of mean changes on a number of fear measures. However, Bergin (1963, 1966) questioned use of only mean change to assess outcome of therapy and has suggested that differential effects of control and treatment procedures may be shown in variances changes, where differences in mean changes are small. This has been labeled the "deterioration effect" to indicate that in therapy some patients get better and some get worse. By contrast, control procedures in

studies cited by Bergin resulted in little change in means or in variance. Treatments also could be compared for effect on variance. That is, two treatments could yield similar mean changes, yet one treatment could have a consistent effect, common to all *Ss*, that would show in a significant variance reduction. To assess such effects in this study, additional analyses were suggested.

The criterion measure, AT, yielded a "pass-fail" score and could not be assessed for variance change. Therefore, AT scores used for ATCS were selected for analyses summarized in Table 4. For CG *Ss* there was no significant change in AT score variance at pre-, post- or follow-up testing. For SDT *Ss*, however, AT score variance showed continuing reduction from pre- to post- ($F = 6.11$, $df = 11/11$, $p < .01$) and from post- to follow-up testing ($F = 13.95$, $df = 11/11$, $p < .01$). The variance change pattern for IT differed from that for both CG and SDT. There was significant *reduction* in variance from pre- to posttesting ($F = 6.43$, $df = 11/11$, $p < .01$) and significant *increase* in variance from post- to follow-up testing ($F = 12.30$, $df = 11/11$, $p < .01$). The pre- and post-AT score variance did not differ ($F = 1.91$, $df = 11/11$, $p > .05$).

TABLE 3

MEAN TRAINING AND TREATMENT TIME, HOURS,
FOR SUBJECTS IN TREATMENT GROUPS
SDT AND IT

Group	Training	Treatment	Total
SDT	3.3	5.8	9.1
IT	1.7	2.4	4.1

Note.—SDT = systematic desensitization therapy; IT = implosive therapy.

DISCUSSION

In this direct comparison of SDT and IT, the therapies were found to be equally effective in reducing snake phobic behavior of otherwise normal adult human beings. Both overt avoidance and verbal reports of fear or anxiety were reduced by learning-theory-derived operations which were directly op-

posed. The therapies did differ in their efficiency since IT was completed in 45% of the time required by SDT. Control procedures resulted in less reduction in phobic behavior. Thus, IT would appear to have been the more efficient of two equally effective methods.

Results of additional analyses, based on Bergin's argument, indicate some necessary qualification of the present findings. These data indicate that the effect of SDT on AT was consistent across Ss and that it continued until the last point of testing. The effect of IT, as assessed by AT, was less consistent. Some Ss improved markedly and maintained their improvement. Others improved on this measure but, at follow-up, showed avoidance behavior that was little different from their pretreatment behavior. Finding a consistent and continuing effect of SDT was consonant both with the findings of Paul (1966, 1967) who used similar individual SDT, and with results of SDT conducted in groups (Paul, 1968; Paul & Shannon, 1966). However, it is difficult to account for differential effects of IT across Ss. Controls were provided for disorder, technique, and therapist variables in this study. The only departure from standard IT procedure was exclusion of scenes specifically suggested by psychodynamic literature, for example, Oedipal scenes. But there is as yet only tentative clinical evidence to indicate that explicit use of psychodynamic themes enhances IT. Further, IT has been shown to be effective where no scenes based on psychodynamic themes were used (Kirchener & Hogan, 1967).

As has been shown in Tables 1 and 2, the major findings of this study concerning effectiveness and efficiency of SDT and IT are consistent with results of previous research (especially Kirchener & Hogan, 1967; Lang et al., 1965). However, some reactions of Ss to the present treatments have not been reported widely in the behavior therapy literature. For example, during SDT, one S showed considerable distress (tears, wringing of hands, tensing) and reported that "wavy lines" that she visualized between scenes had turned into moving, frightening snakes. Similarly, an SDT S began sobbing during visualization of a

TABLE 4
MEANS AND VARIANCES OF AVOIDANCE TEST SCORES
AT PRE-, POST-, AND FOLLOW-UP TESTING

Testing	CG		SDT		IT	
	M	Variance	M	Variance	M	Variance
Pre	10.08	18.81	8.42	17.90	8.16	14.33
Post	9.00	27.71	1.75	2.93	1.91	2.23
Follow-up	7.90	27.54	1.25	.21	3.16	27.42

Note.—CG = control group; SDT = systematic desensitization therapy; IT = implosive therapy.

scene in which she previously was comfortable. She reported that the scene, which included a snake in a closed cage, had developed suddenly and unexpectedly into a frightening one in which the snake came toward her. These may be examples of the same phenomenon reported by Weitzman (1967) as "flow of visual imagery" and by Weinberg and Zaslove (1963) as "resistance." However, the degree of disturbance that Ss in this study experienced has not been reported elsewhere.

Similar reactions to scenes during IT, of course, were expected. As required by IT, instructions for S to continue imagining frightening scenes were given when Ss showed disturbance and, after a time, S ceased to show disturbance. However, two Ss in the IT group showed a disturbance between IT sessions. One S, a college student, reported that after therapy she consistently visualized snakes when she shut her eyes, that her sleep was disturbed, that she could not attend lectures, and that she was unwilling to participate in the second session of IT. She was persuaded to continue in IT, and, at post- and follow-up testing, held the snake. A similar report of being very upset was given by a nurse who said that her work was disrupted frequently by her crying when she thought of snakes. She too participated in the second IT session and was successful at post- and follow-up testing. Similar between-treatment effects of IT have been reported by Roger Kotila.³ For example, one of Kotila's Ss began compulsively to check her bed for snakes every night. Results such as these underline the criticisms of behavior therapy (Breger & McGaugh, 1965; Weitzman, 1967) by indicating that the

³ Roger Kotila, personal communication, May 21, 1968.

operation of instructing S to imagine scenes does not lead necessarily to reliable responses.

Findings of the present study support the view that IT is an effective, efficient behavior modification technique for use with phobias. Further, the results make more explicit evidence for SDT as an effective technique that has a consistent and continuing effect across Ss and across time. However, disturbances reported by both SDT and IT Ss during imagery, again raise the serious question of what Ss actually do following instructions to "imagine a scene."

REFERENCES

- BARRETT, C. L. Systematic desensitization therapy (SDT) versus implosive therapy (IT): A comparative study of two methods of reducing the snake phobic behavior of otherwise normal adults. Unpublished doctoral dissertation, University of Louisville, 1967.
- BERGIN, A. E. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, 1963, 10, 244-250.
- BERGIN, A. E. Some implications of psychotherapy research for therapeutic practice. In G. E. Stollak, B. G. Guernsey, & M. Rothberg (Eds.), *Psychotherapy research*. Chicago: McNally, 1966.
- BREGER, L., & MCGAUGHER, J. L. Critique and reformulation of "learning theory" approaches to psychotherapy and neurosis. *Psychological Bulletin*, 1965, 63, 338-358.
- ENDLER, N. S., HUNT, J. McV., & ROSENSTEIN, A. J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, 76(17, Whole No. 536).
- HOGAN, R. A. Implosive therapy in the short term treatment of phobias. *Psychotherapy, Theory, Research and Practice*, 1966, 3, 25-31.
- KIRCHENER, J. H., & HOGAN, R. A. The therapist variable in the implosion of phobias. *Psychotherapy, Theory, Research and Practice*, 1967, 4, 102-104.
- LANG, P. J., LAZOVIC, A. D., & REYNOLDS, D. J. Desensitization, suggestibility and pseudotherapy. *Journal of Abnormal and Social Psychology*, 1965, 70, 395-402.
- LANG, P. J., & LAZOVIC, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology*, 1963, 66, 519-525.
- LEVIS, D. J., & CARRERA, R. Effects of ten hours of implosive therapy in the treatment of outpatients: A preliminary report. *Journal of Abnormal Psychology*, 1967, 72, 504-508.
- PAUL, G. L. *Insight versus desensitization in psychotherapy*. Stanford: Stanford University Press, 1966.
- PAUL, G. L. Insight versus desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology*, 1967, 31, 333-348.
- PAUL, G. L. Two years follow-up of systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, 1968, 73, 119-130.
- PAUL, G. L., & SHANNON, D. T. Treatment of anxiety through systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, 1966, 71, 124-135.
- SOLOMON, R. L., KAMIN, L. J., & WYNN, L. C. Traumatic avoidance learning: The outcomes of several extinction procedures with dogs. *Journal of Abnormal and Social Psychology*, 1953, 48, 291-302.
- STAMPFL, T. G. Implosive therapy: A learning-theory-derived psychodynamic therapeutic technique. Unpublished manuscript, University of Illinois, 1961.
- STAMPFL, T. G., & LEVIS, D. J. Essentials of implosive therapy: A learning-theory-based psychodynamic behavioral therapy. *Journal of Abnormal Psychology*, 1967, 72, 496-503.
- WALK, R. D. Self ratings of fear in a fear-invoking situation. *Journal of Abnormal and Social Psychology*, 1956, 52, 171-178.
- WEINBERG, N. H., & ZASLOVE, M. "Resistance" to systematic desensitization of phobias. *Journal of Clinical Psychology*, 1963, 19, 179-181.
- WEITZMAN, B. A. Behavior therapy and psychotherapy. *Psychological Review*, 1967, 74, 300-317.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.
- WOLPE, J., & LANG, P. J. A fear survey schedule for use in behavior therapy. *Behavior Research and Therapy*, 1964, 2, 27-30.

(Received July 15, 1968)

SPAN OF APPREHENSION IN ACUTE SCHIZOPHRENICS

JOHN M. NEALE, CURTIS W. McINTYRE,¹ ROBERT FOX, AND RUE L. CROMWELL

Vanderbilt University

The span of apprehension for two schizophrenic groups (good premorbid paranoid and poor premorbid nonparanoid) was compared with the span for a normal control group (hospital aides) by a forced-choice letter-recognition method analogous to signal-detection methodology. This procedure, in conjunction with a quantitative model of the span, provides an indirect but relevant measure insensitive to variance imposed by nonperceptual systems. The span size was the same for all three groups when visual "noise" was absent. In the presence of noise, span size for both schizophrenic groups was about 50% smaller than that of the control group. It is argued that the reduction in span represents a true deficit in attention, uncontaminated by extraneous variables.

A considerable portion of the literature on schizophrenia (Bleuler, 1951; Silverman, 1964; Venables, 1964) suggests that a central characteristic of the schizophrenic process involves a deficit in attention. Although the hypothesis is certainly reasonable, it is difficult to evaluate empirically (e.g., Neale & Cromwell, 1969). One problem is posed by vagaries in the definition of attention. The difficulties are further compounded by the special psychophysical measurement problems posed by schizophrenics themselves. Because of their labile motivational systems and often retarded response systems, it is essential that care be taken so that aberrations in these systems do not interfere with the measurement of attention.

In this article an experiment is described that attempts to circumvent these difficulties by employing an explicit model of one aspect of attention and by employing a measurement technique that is relatively insensitive to perturbations imposed by extraneous systems. Specifically, the span of apprehension in schizophrenics was measured. The span of apprehension is a measure of the amount of information a person can simultaneously attend. Span was measured by a method derived from signal-detection procedures. The span of schizophrenics was found to be less than that of control Ss, and we argue that this represents a demonstration of a deficit in attention that

is uncontaminated by extraneous nonperceptual variables.

The span of apprehension is a measure of considerable relevance to any theory of attention. In the earliest experiments on the span, stimuli such as letters of the alphabet were exposed so briefly that eye movements could not occur (i.e., tachistoscopic presentation). The Ss were requested merely to report the letters they had seen. Typically, Ss correctly reported four or five letters even though many more were presented initially. But it has now been demonstrated that the limitation to four or five letters is one imposed by memory rather than by the perceptual capacity of the Ss (Sperling, 1960). Sperling showed that the span of apprehension is a positive function of the amount of information in the display by using a reporting method that removed much of the burden from S's immediate memory.

Current theoretical conceptions of the span of apprehension involve the idea that all of the information in the display is available for processing for several hundred milliseconds after the display terminates. A hypothetical central scanning device is assumed to examine the elements in the display serially. If the scanner attends to an element, that element passes on to a subsequent information-processing stage, but if the excitation from an element dissipates before the scanning device can attend to it, the information associated with the excitation is lost.

Quite recently Estes (1965) has developed an explicit quantitative model of the span of apprehension that specifies an indirect method

¹ Requests for reprints should be sent to Curtis W. McIntyre, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

TABLE 1
DESCRIPTION OF SUBJECT GROUPS

Group	Age	Education
Normals		
<i>M</i>	29.83	11.23
<i>SD</i>	8.64	2.79
Good paranoids		
<i>M</i>	33.24	8.93
<i>SD</i>	5.95	1.65
Poor nonparanoids		
<i>M</i>	29.07	9.00
<i>SD</i>	8.90	2.06

for estimating the span. Although the model is not complex, there is no need to review its formal assumptions here. It will be helpful, however, to state verbally the general rationale of the measurement procedure. From the standpoint of the *S* the basic task is simply a forced-choice letter-recognition task, where on any tachistoscopic exposure either one of two letters will appear. The *S* knows what these letters are, say for example the Letter T and the Letter F, and he also knows the a priori probability of their occurrence. The letters can be thought of as analogous to the signals one would be required to detect in a standard signal-detection experiment (see Green & Swets, 1966). If the signal or target letters are placed in an array of other irrelevant letters, then it is intuitively clear that these irrelevant letters can be considered analogous to noise in the signal-detection sense. That is, if an *S* is going to make above-chance recognitions of the signal letters, his scanning process must discriminate the signal letter from the irrelevant noise letters. The number of letters that the scanner examines, of course, is an operational statement of the span of apprehension.

Estes has shown how the number of letters scanned can be estimated from the percentage of correct recognitions and from the number of letters physically present in a display. He has demonstrated that as the number of letters in a display increases the probability of correct recognition predictably declines, but the number of elements scanned or processed, that is, the span of apprehension, increases with display size.

It should be emphasized that the *S*'s task remains the same, independent of the num-

ber of noise letters in a display. He merely gives a forced-choice recognition response; for example, it was a T or it was an F. Subjectively, the *S*'s task seems similar regardless of the number of noise elements presented in the display. Note that the scanning process presumed to operate does not involve eye movements; rather, it is a central process.

In the present experiment the span of schizophrenics and of control *Ss* under two information conditions was measured. In one condition no noise letters were presented. In the second condition, the signal letter was surrounded by seven noise letters. Explicit a priori predictions were not made. While constructs such as narrowed attention (Venables, 1964) and extensive scanning (Silverman, 1964) have been applied to the performance of schizophrenics on various tasks, the unique operations associated with these constructs suggests that it would not be useful to relate them directly to this experiment.

METHOD

Subjects

Forty-five males were selected to form three groups: 15 good premorbid paranoid schizophrenics; 15 poor premorbid nonparanoid schizophrenics; and 15 hospital aides. The groups were equated for age and level of education. See Table 1 for group means and standard deviations. Differences between these means were not significant. All patient *Ss* were hospitalized for less than 3 yr. and most were on phenothiazine medication. Premorbid adjustment status conformed to the criteria suggested by Held and Cromwell (1967), while paranoid classification was based upon both clinical history and psychiatric diagnosis.

Apparatus

The stimulus displays, consisting of arrays of letters, were designed and constructed in the following way. Each array was made by typing letters on a white index card, which in turn was mounted on a cardboard insert that facilitated rapid insertion and removal from a tachistoscope. To allocate letters in each array, an imaginary matrix of 16 letter spaces was located at the center of each card. Each matrix (4 letters wide and 4 letters high) subtended $2^\circ \times 3^\circ$ of visual angle. Each array contained one of two target letters (T or F); each letter appeared once on two separate cards in each of the 16 possible matrix positions, yielding a total of 32 T arrays and 32 F arrays. One set of 32 arrays, composed of 16 T cards and 16 F cards, contained only the target letter (Matrix Size 1). A second set of 32 arrays, 16 T cards and 16 F cards, contained the target letter plus

7 additional noise letters (Matrix Size 8). In this second set of stimulus displays, a target letter was first located at 1 of the 16 possible positions and then 7 different consonant letters were randomly selected without replacement and randomly allocated to 7 of the 15 remaining positions in the matrix. Each letter in the matrix subtended approximately $12' \times 18'$ of visual angle, while letters occurring in adjacent cells were separated vertically by $30'$ of visual angle and laterally by $24'$ of visual angle.

The stimulus arrays were exposed in a two-channel tachistoscope (Scientific Prototype, Model 800-E). One field, the constantly illuminated fixation field, contained four black lines $12'$ in width arranged to form a $2\frac{1}{2}^\circ \times 3\frac{1}{2}^\circ$ cornerless rectangle. The center of the rectangle defined the fixation area. The fixation field luminance was 9.6 ftl. The stimulus displays were exposed for 90 msec. in the other field of the tachistoscope. The exposure field luminance was 78 ftl.

Procedure

Each *S* was given a standard set of verbal instructions followed by 20 practice trials. Practice trials were given to habituate *S* to the task and to insure that *S* understood the instructions. The *Ss* failing to perform above chance on the practice trials were to be dropped from the study; only one *S*, from the good premorbid paranoid group, was eliminated for this reason. During the practice trials, *S* was presented with two randomly selected blocks of 10 trials, each block drawn from a separate matrix size with the restriction of an equal occurrence of the critical elements (5 Ts, 5 Fs). After the practice trials each *S* was allowed to rest for 2 min. before the experimental trial blocks began. Each *S* then viewed two sets (Matrix Sizes 1 and 8) of four 10-trial blocks, each separated by a 2-min. rest period. Order of presentation of the matrix sizes was controlled by an ABBA, BAAB counterbalancing scheme within *Ss*. Each *S* received a different randomization of stimulus materials. Intertrial interval varied with the individual *S*'s response rate, but an average of approximately one trial every 15 sec. was maintained. A 30-sec. rest was given after each block of 10 trials.

RESULTS AND DISCUSSION

The basic data were the mean number of correct detections by each of the three groups for the two matrix sizes. These data are plotted in Figure 1. An analysis of variance revealed significant effects for groups ($F = 3.28$, $df = 2/42$, $p < .05$); for matrix size ($F = 598.04$, $df = 1/42$, $p < .001$), and for Groups \times Matrix Size ($F = 5.56$, $df = 2/42$, $p < .01$). To determine the source of the interaction one-way analyses were performed. For Matrix Size 1 the groups did not differ

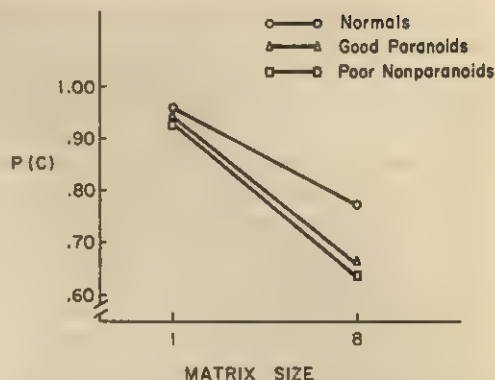


FIG. 1. Probability of correct response as a function of matrix size for the control and schizophrenic groups.

($F = 2.20$, $df = 2/42$, $p > .10$); for Matrix Size 8 the groups did differ significantly ($F = 58.16$, $df = 2/42$, $p < .001$). A Newman-Keuls test revealed that at Matrix Size 8 both schizophrenic groups performed at a significantly lower level than did the control group ($p = .01$) and, further, the schizophrenic groups did not differ between themselves.

The mean detection scores for each of the three groups at Matrix Size 8 were used to compute the estimated span of apprehension in accord with the Estes model. By that analysis the control *Ss* processed 4.48 letters while the good premorbid paranoid group processed 2.52 letters and the poor premorbid nonparanoid group processed 2.40 letters. Thus it appears that the span of apprehension for schizophrenics is approximately one-half that of the control *Ss*.

It is unlikely that the poor performance of the schizophrenics was a function of their age, education, or level of drug dosage,² for none of the correlations between these variables and detection scores was significant ($r = -.06, -.02, +.05$, respectively).

Further, it is unlikely that the poor performance of the schizophrenics at Matrix Size 8 was due to a general failure to comprehend the task or a similar kind of generalized deficit, since the schizophrenics clearly performed at a level comparable to that of the control group under the Matrix Size 1 condition.

²The authors would like to thank J. D. Griffith for his assistance in ranking drug levels.

Therefore, we are forced to conclude that the relatively poor performance of the schizophrenic group under the Matrix Size 8 condition results from an inability to process efficiently the information contained in the eight-letter display. In brief, the schizophrenics have a reduced span of apprehension. This reduction in span seems to be attributable to some inefficiency of the central processing mechanism that analyzes and encodes the information presented in the briefly exposed display. This interpretation seems most reasonable because many of the potential alternative explanations based upon the operation of extraneous variables have been minimized by the design and procedures employed in this experiment.

With respect to the nature of the deficiency that produces a reduction in the span of apprehension, several hypotheses can be considered. For example, considerable evidence suggests that two primary characteristics of schizophrenics are reduction in speed of responding and greater susceptibility to distraction. These concepts of "slowness" and "distractibility" are induced from experimental situations quite different from the present experiment and may not be applicable to tachistoscopic recognition. Yet it is possible to entertain the hypothesis that the noise letters act as potent distractors that impede the search for the target letter. Similarly, it is possible that processing of information from a brief display simply proceeds at

a slower rate in schizophrenics. Either distractibility or slowness would result in a reduced span of apprehension. Decisions between these alternatives and others can be made through further research using the design of the present experiment and the analytic techniques of the span of apprehension model.

REFERENCES

- BLEULER, E. The basic symptoms of schizophrenia. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia University Press, 1951.
- ESTES, W. K. A technique for assessing variability of perceptual span. *Proceedings of the National Academy of Sciences*, 1965, 54, 403-407.
- GREEN, D. M., & SWETS, J. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HELD, J. M., & CROMWELL, R. L. Premorbid adjustment in schizophrenia: An evaluation of a method and some general comments. Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, April 1967.
- NEALE, J. M., & CROMWELL, R. L. Attention and schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 5. New York: Academic Press, 1969, in press.
- SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, 71, 353-379.
- SPELTING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, 74(11, Whole No. 493).
- VENABLES, P. H. Input dysfunction in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 1. New York: Academic Press, 1964.

(Received July 16, 1968)

EXPERIMENTAL CONTROL OF SEXUAL DEVIATION THROUGH MANIPULATION OF THE NOXIOUS SCENE IN COVERT SENSITIZATION¹

DAVID H. BARLOW,² HAROLD LEITENBERG, AND W. STEWART AGRAS³

University of Vermont

Pedophilic behavior in one *S* and homosexual behavior in another *S* were decreased, increased, and once again decreased by introducing, removing, and reintroducing the noxious scene in covert sensitization, a form of aversion therapy. The results indicated that verbal description of a nauseous scene was an effective aversive stimulus, and that pairing this scene with scenes of the undesired behavior was responsible for declines in deviant sexual behavior during covert sensitization.

The treatment of sexual deviation by aversion therapy has recently become more popular (Feldman, 1966). All studies reported to date, however, are either uncontrolled single case reports or uncontrolled group outcome studies. It is not possible to determine from such data whether the reported success is due to the conditioning technique or to another of the myriad of variables present either singly or in combination in the therapeutic situation.

The usual way to isolate relevant treatment variables is the group comparison in which a control group experiences the same treatment as an experimental group with the exception of the independent variable. Group differences are then ascribed to the independent variable. An alternative method used in this study is to demonstrate experimental control in individual cases. One way of doing this is to sequentially vary some aspect of treatment while measuring associated changes in a well-specified, clinically relevant behavior (cf. Agras, Leitenberg, & Barlow, 1968; Leitenberg, Agras, Thomson, & Wright, 1968; Wolf, Birnbrauer, Williams, & Lawler, 1965).

¹This study was supported in part by United States Public Health Service Clinical Research Center Grant FR-109 and National Institute of Mental Health Grant MH-13651. Part of the data was presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 1968.

²Requests for reprints should be sent to David H. Barlow who is now at the Department of Psychiatry, University of Mississippi Medical Center, Jackson, Mississippi 39216.

³Now at University of Mississippi Medical Center.

In this experiment the aversion therapy studied was covert sensitization (Cautela, 1966). In this procedure, descriptions of extremely noxious scenes are paired with scenes of the undesired behavior. Uncontrolled case studies suggest that covert sensitization may modify sexual deviation. Cautela (1967) treated two homosexuals with this method resulting in reports by the patients of decreases in homosexual behavior, while Davison (1968) treated a case of long-standing sadistic fantasies with a combination of covert sensitization and positive conditioning to heterosexual stimuli and achieved a remission. Other case studies include Kolvin's (1967) successful treatment of a fetish.

The crucial procedure in covert sensitization is thought to be the pairing of verbal descriptions of a noxious scene with descriptions of scenes involving the undesired behavior. This notion was tested in the present study in two cases of sexual deviation by first pairing the two scenes, then removing the noxious scene while holding other variables constant, and finally pairing the scenes once more. If the pairing procedure is critical to therapeutic success, then any improvement should be reversed or stopped when the noxious scene is removed.

METHOD

Subjects. The *S*₁ was a 25-yr.-old married male who reported a 13-yr. history of pedophilic experiences ranging from fantasies to several instances of sexual contact. He came for treatment after the solicitation of a neighbor's 9-yr.-old daughter was discovered. The neighbor, rather than report him to

the police, insisted that he seek psychiatric help. Sexual relations with his wife averaged twice a month.

The S_2 was a 32-yr.-old married male who reported a 14-yr. history of homosexual experiences averaging about three contacts per week, usually in public toilets. He recently had fallen in love with a "boyfriend," which was threatening his marriage and which motivated him to seek treatment. Sexual relations with his wife, although prevalent early in the marriage, had been virtually nonexistent for the previous 3 yr.

Measures. A hierarchy of sexually arousing scenes was constructed. For S_1 the hierarchy contained 45 scenes involving small girls. For example, the top item in the hierarchy was, "You are alone in a room with a very sexy looking 10-year-old girl with long blond hair." One of the least arousing items was, "While driving in your car you can see a small, thin, six-year-old girl walking down the street." For S_2 the hierarchy contained 27 scenes involving males. The top scene in this hierarchy pictured a "30-year-old, well-built, good-looking fellow, dressed neatly and standing in a large rest room." At the other end of the hierarchy was "an older, sloppily-dressed fellow" in the same location. Five scenes involving the boyfriend were added to these 27 items. The five scenes ranged from sexual contact to talking with him on the telephone. Thus, the total number of items was 32.

Base-line procedures were then instituted for five sessions, during which three measures were taken: First, S_s kept a small notebook in which they recorded each time that they were sexually aroused, for S_1 by the sight of an immature girl, and for S_2 by the sight of a mature male. The S_s , as an outpatient, recorded these events daily. Due to the distance of his home from the treatment center, S_2 remained in the hospital during the week and returned home on weekends. He was, therefore, instructed to visit one of the local bars 4 nights a week for 1-2 hr., during which time he recorded incidents of sexual arousal.

Second, the hierarchy scenes were typed on individual cards. For S_1 the 45 cards were divided into five packs of 9 cards each, chosen in such a way that each pack contained an equally arousing set of scenes. The 32 cards of S_2 were not divided. The S_s then was given one of the packs, while S_2 was given the whole set. On every experimental day each S was asked to enter a separate room where he could sort each of the cards into one of five envelopes marked 0-4. He was told,

The numbers on the envelopes represent amount of sexual arousal, 0 equals no arousal, 1 equals a little arousal, 2 a fair amount, 3 much, and 4 very much arousal. I would like you to read the description of each scene and place a card in the envelope which comes closest to how arousing the scene is to you at this moment.

If he then put one card in the Number 4 envelope and one in the Number 3 envelope, his score to that point would be 7.

Third, galvanic skin responses (GSRs) were re-

corded from a Grass polygraph (Model 7) with silver, silver-chloride, palmar and wrist electrodes to six selected scenes from the hierarchy of each S . The same six scenes were used throughout the experiment. The S was relaxed and then asked to close his eyes and imagine the scene as clearly as possible. GSR deflection was measured as change in log conductance that was averaged for the six scenes. This measure was taken during every session of the base line and before every second session during treatment.

Procedure. During the initial interview S_s were told,

We view this type of problem as a bad habit that has been picked up over the years and we are going to try to break that habit. We will be using some techniques that have been found effective in dealing with bad habits such as this one. However, no matter how successful, every treatment has its ups and downs so don't get discouraged.

This last sentence was designed to allay any fear of ultimate failure that might have arisen when extinction procedures were later introduced. The S_s were then trained in deep muscle relaxation following the method described by Jacobson (1938). The highest six items in S_1 's hierarchy were chosen for sensitization. The S_1 was seen twice a week as an outpatient. Since S_2 's major problem at the outset was an "overwhelming" attraction for his boyfriend, the five scenes describing the boyfriend were chosen for sensitization along with one scene concerning sexual contact with "pickups" in public toilets. The S_2 was seen 5 days a week as an inpatient, and from Experimental Day 12 on was generally seen twice a day.

In each session S_s were given relaxation instructions and presented with eight scenes. In four scenes S was described approaching the small girl (male), feeling nauseous and vomiting. For example, in one of the homosexual scenes, S_2 was described approaching his boyfriend's apartment.

As you get closer to the door you notice a queasy feeling in the pit of the stomach. You open the door and see Bill lying on the bed naked and you can sense that puke is filling up your stomach and forcing its way up to your throat. You walk over to Bill and you can see him clearly, as you reach out for him you can taste the puke, bitter and sticky and acidic on your tongue, you start gagging and retching and chunks of vomit are coming out of your mouth and nose, dropping onto your shirt and all over Bill's skin.

The description of the nauseous scene was usually expanded and lasted from 30 to 60 sec. In the remaining four scenes S would be described approaching the small girl (male) and beginning to feel nauseous. At that point he would turn, start walking away from the scene, and immediately feel relieved and relaxed. The scenes were presented randomly. The sexually arousing scene was presented for approximately 10 sec., the nauseous scene for 30-60 sec., with an intertrial interval of 30 sec.

After 6 acquisition sessions consisting of 48 pairings for S_1 , and 13 acquisition sessions consisting of 104 pairings for S_2 , extinction was introduced. The

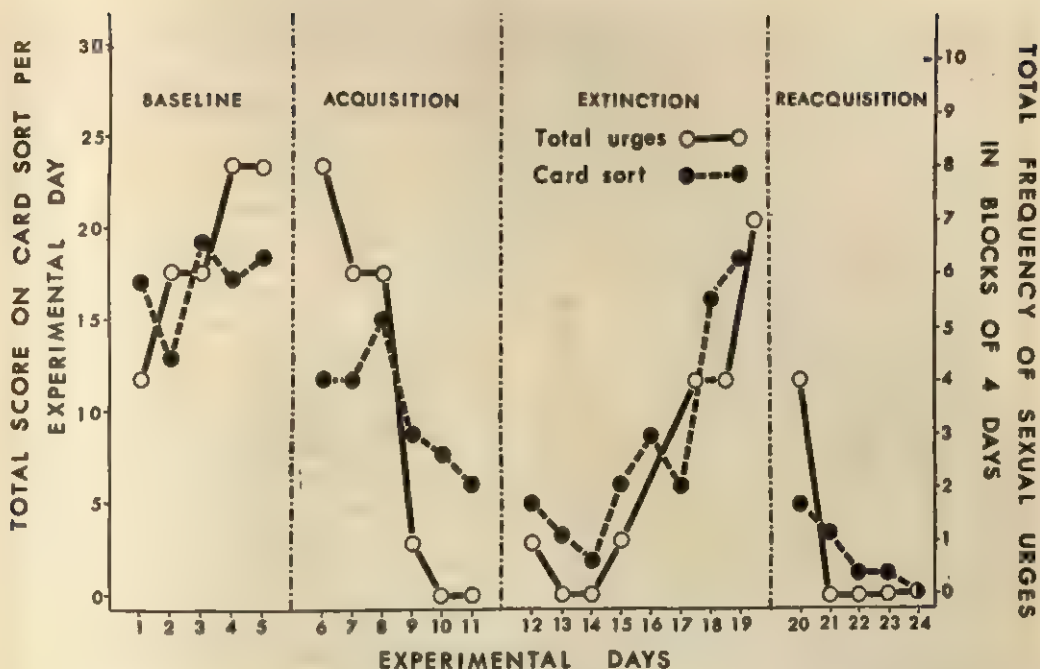


FIG. 1. Total score on card sort per experimental day and total frequency of pedophilic sexual urges in blocks of 4 days surrounding each experimental day. (Lower scores indicate less sexual arousal.)

S was told, "For the next few sessions we are going to change the procedure a bit. We have found that this is the best course of action at this time. Remember to imagine only what I describe." Note that instructions during extinction suggest continued progress. The procedure during extinction consisted of presenting the sexually arousing scene for 10 sec., leaving the 30-sec. nauseous interval blank, and saying, "Stop imagining that," at the end of the interval. All other therapist behaviors remained constant.

After eight extinction sessions consisting of 64 scene presentations for each S, reacquisition was introduced. The design, then, consists of base-line measures; acquisition, in which sexually arousing scenes are paired with nauseous scenes; extinction, in which sexually arousing scenes are presented alone; and reacquisition.

RESULTS

Figure 1 plots the total score of the card sort for each experimental day and the total frequency of sexual arousal in blocks of 4 days surrounding each experimental day for S₁. These 4 days consisted of the 2 days before the session, the day of the session, and the day after the session.

During the base-line phase there was no treatment. The card sort remained stable during this period while reports of pedophilic urges steadily rose. Acquisition, in which the

nauseous scene was paired with the sexually arousing scene, resulted in a sharp drop in both measures of inappropriate sexual arousal. In extinction the nauseous scene was omitted, resulting in an increase in both measures of sexual arousal.

The instruction preceding this extinction phase conveyed an expectancy of therapeutic progress. This expectancy was verified when S said, toward the end of extinction, "I know you're doing your best, but I guess I'm just not cut out for this treatment." At this time S₁ became upset and depressed.

In reacquisition, the nauseous scene was reintroduced and measures of sexual arousal dropped to zero. During this phase S₁ reported increased heterosexual behavior.

GSR, analyzed as change in log conductance, was averaged over each of the four experimental periods. The means (*M*s) and standard deviations (*SD*s) in μ mhos are: base line, *M* = 221 and *SD* = 141; acquisition, *M* = 27 and *SD* = 29; extinction, *M* = 200 and *SD* = 92; reacquisition, *M* = 30 and *SD* = 23.

Although there is a great deal of variability, it is clear that S₁'s arousal rose during extinction and declined during reacquisition.

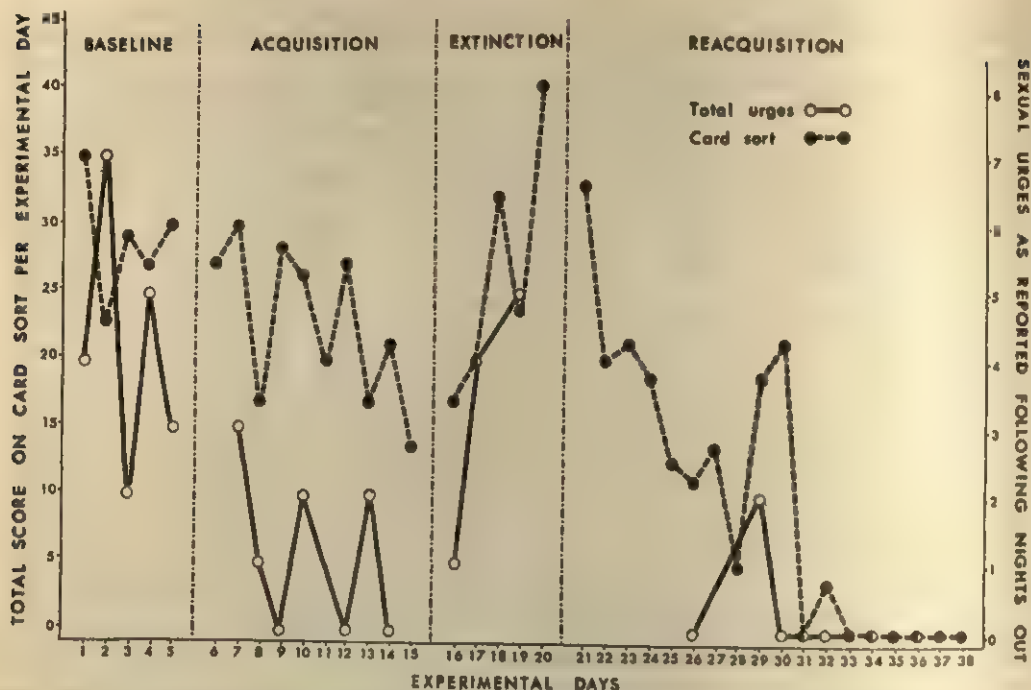


FIG. 2. Total score on card sort per experimental day and homosexual urges as reported following nights out. (Lower scores indicate less sexual arousal.)

Figure 2 plots the total score of the card sort on each experimental day for S_2 and also the frequency of sexual arousal on each night out. For various reasons S_2 was not able to go out each day that sessions were held. At one point an illness prevented him from visiting the local bars for 4 consecutive nights (Experimental Days 22–25). Therefore, points referring to frequency of sexual arousal are plotted only on experimental days for which there are data.

During the base-line phase both the card sort and reports of homosexual urges were relatively stable. The acquisition phase resulted in a drop in both measures of sexual arousal. In the course of extinction both measures rose sharply with the total score of the card sort rising above base-line levels.

During this phase S_2 also became depressed, and near the end of the period said: "This treatment isn't doing much good. I'm just as bad now as I was when I came in." At this time he engaged in his first homosexual affair since treatment began. Due to an illness of the therapist, there was a 9-day period be-

tween Experimental Days 20–22 during which S was only seen once (Day 21). Throughout this period S_2 generally refused to leave the hospital, even for the weekend, and remained quite depressed.

In reacquisition, measures of sexual arousal dropped to zero. All sessions beyond the twentieth session (Day 31) were self-administered. At this time, S_2 also reported increases in heterosexual behavior and acquisition of heterosexual fantasies.

Although mean magnitude of GSR declined sharply during acquisition, it did not recover during extinction and therefore could not be interpreted.

DISCUSSION

These findings demonstrate that pairing a noxious scene with a sexually arousing scene is a crucial procedure in covert sensitization. The effect of this pairing was separated from other psychotherapeutic variables such as therapeutic instructions, patient expectancies of improvement, and rapport between patient and therapist since these variables were still

present during extinction when the noxious scene was removed, and yet behavior regressed.

The two noxious stimuli most often used in aversion therapy are emetic drugs (James, 1962) and more recently shock (Feldman & MacCulloch, 1965). This experiment demonstrates that an intensely imagined noxious scene can also act as an effective aversive stimulus. There would seem to be several advantages to using a noxious scene in clinical situations. First, the patient is less likely to refuse treatment because of the pain involved, (Rachman, 1965). Second, it can be widely employed by many therapists, since it does not require drugs or apparatus.

Generally, there have been two therapeutic strategies in dealing with sexual deviations: decreasing deviant sexual behavior and increasing appropriate heterosexual behavior. This experiment indicates that covert sensitization alone is sufficient to decrease deviant behavior. More controlled research employing precise measures of heterosexuality is needed to determine what treatment variables are responsible for increases in heterosexual behavior. Further research is also necessary to ascertain whether reductions in deviant behavior resulting from covert sensitization would be maintained without continued treatment and careful attention to later sexual adjustment.

The present study does not completely overcome a measurement problem common to most psychotherapy research, namely, an excessive reliance on the patient's subjective reports of progress. Although there is evidence that GSR is a valid index of sexual arousal (Solyom & Beck, 1967; Wenger, Averill, & Smith, 1968), the remaining measures are not behavioral in the sense of being publicly observable. While observable and objective behavioral measures have been devised for some neurotic (e.g., Leitenberg et al., 1968) and psychotic disorders (Ayllon & Azrin, 1965), it is more difficult in cases of sexual deviation where more than one person is involved. Recent studies in our laboratory and elsewhere (McConaghy, 1967), however, suggest that changes in penile volume to slides of inappropriate sexual objects, a measure originally devised by Freund (1963), may be useful in the measurement of sexual deviation.

REFERENCES

- AGRAS, W. S., LEITENBERG, H., & BARLOW, D. H. Social reinforcement in the modification of agoraphobia. *Archives of General Psychiatry*, 1968, 19, 423-427.
- AYLLON, T., & AZRIN, N. H. The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, 1965, 8, 357-383.
- CAUTELA, J. R. Treatment of compulsive behavior by covert sensitization. *Psychological Record*, 1966, 16, 33-41.
- CAUTELA, J. R. Covert sensitization. *Psychological Reports*, 1967, 20, 459-468.
- DAVISON, G. C. The elimination of a sadistic fantasy by a client-controlled counterconditioning technique: A case study. *Journal of Abnormal Psychology*, 1968, 73, 84-90.
- FELDMAN, M. P. Aversion therapy for sexual deviations: A critical review. *Psychological Bulletin*, 1966, 65, 65-79.
- FELDMAN, M. P., & MACCULLOCH, M. J. The application of anticipatory avoidance learning to the treatment of homosexuality. I. Theory, technique and preliminary results. *Behavior Research and Therapy*, 1965, 2, 165-183.
- FREUND, K. A laboratory method for diagnosing predominance of homo- or hetero-erotic interest in the male. *Behavior Research and Therapy*, 1963, 1, 85-93.
- JACOBSON, E. *Progressive relaxation*. Chicago: University of Chicago Press, 1938.
- JAMES, B. A case of homosexuality treated by aversion therapy. *British Medical Journal*, 1962, 1, 768-770.
- KOLVIN, I. "Aversive Imagery" treatment in adolescents. *Behavior Research and Therapy*, 1967, 5, 245-248.
- LEITENBERG, H., AGRAS, W. S., THOMSON, L. E., & WRIGHT, D. E. Feedback in behavior modification: An experimental analysis in two phobic cases. *Journal of Applied Behavior Analysis*, 1968, 1, 131-137.
- MCCONAGHY, N. Penile volume change to moving pictures of male and female nudes in heterosexual and homosexual males. *Behavior Research and Therapy*, 1967, 5, 43-48.
- RACHMAN, S. Aversion therapy: Chemical or electrical? *Behavior Research and Therapy*, 1965, 2, 289-299.
- SOLYOM, L., & BECK, P. R. GSR assessment of aberrant sexual behavior. *International Journal of Neuropsychiatry*, 1967, 3, 52-59.
- WENGER, M. A., AVERILL, J. A., & SMITH, D. D. B. Autonomic activity during sexual arousal. *Psychophysiology*, 1968, 4, 468-478.
- WOLF, M. M., BIRNBAUER, J. S., WILLIAMS, T., & LAWLER, J. A note on apparent extinction of vomiting behavior of a retarded child. In L. P. Ullman & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart & Winston, 1965.

(Received July 26, 1968)

ANXIETY AND OVERGENERALIZATION: NEGATIVE RESULTS¹

E. NEIL MURRAY²

State University of New York at Buffalo

The effect of chronic and experimental anxiety on stimulus generalization was investigated. The paradigm consisted of conditioning a voluntary motor response to lines 8.9 cm. long by means of verbal reinforcement and recording amplitude of generalized response to six shorter lines as well as to the training stimulus. No relationship was found between either chronic or experimental anxiety and the *height* of the gradients for men or for women. For *slope*, a significant difference was obtained only with experimental anxiety and only for men, but this result failed to replicate. Conclusions are: (a) chronic anxiety level is unrelated to stimulus generalization; (b) experimentally induced anxiety has not been demonstrated convincingly as having an effect on stimulus generalization; (c) the popular notion that anxiety leads to overgeneralization is based on the weakest of experimental evidence.

The notion that increased anxiety leads to overgeneralization or increased stimulus generalization is often cited as an example of the application of general principles of learning to the understanding of symptoms of emotional disturbance. According to Mednick (1958), broadened generalization causes more distant stimuli to become associated with anxiety, and anxiety level mounts accordingly, thereby establishing a positive feedback system of ever-increasing anxiety.

Although a study by Mednick (1957) is widely quoted as favoring the anxiety-over-generalization hypothesis, greater generalization in a middle-anxious group and no difference between high- and low-anxious groups was actually found, and these results cannot be accepted as positive evidence. In fact, there is only one investigation of the effects of anxiety on stimulus generalization that has yielded results that clearly support the hypothesis (Rosenbaum, 1953, 1956). Rosen-

baum obtained gradients from psychiatric patients who had been rated high and low in clinical anxiety and gradients from college students who both scored high and low on the Taylor Manifest Anxiety (*MA*) scale (Taylor, 1953) and were subjected to one of three induced experimental anxiety conditions: strong shock, weak shock, or buzzer. He found that students in the strong shock group produced a higher and flatter gradient than those in the weak shock and buzzer groups, and that within the strong shock group students with high *MA* scale scores produced a higher gradient than those with low *MA* scale scores.

These results must be evaluated cautiously for there were several methodological restrictions concerning the definition and measurement of stimulus generalization in Rosenbaum's paradigm: (a) *Ss* were instructed, rather than conditioned, to respond to a given "correct" stimulus and not to respond to other stimuli; (b) *Ss* were instructed to respond as rapidly as possible upon presentation of the stimuli; (c) generalization stimuli were introduced during training trials, and large numbers of additional training trials ("booster" trials) were run during the generalization series. These procedures define generalization in terms of error tendencies, and they also resemble the operations employed in studies of discrimination learning more closely than those traditionally used in

¹ The investigation was supported in part by United States Public Health Service Grant MH-07080. This article is based on a doctoral dissertation submitted to the University of Pittsburgh. The author wishes to thank Arnold H. Buss, chairman of the committee, for his advice and direction at every stage of the research. For their thoughtful comments, appreciation is expressed also to A. David Lazovic, William J. Meyer, and Arthur R. Orgel, members of the committee.

² Requests for reprints should be sent to E. Neil Murray, Department of Psychology, State University of New York, Buffalo, New York 14214.

studies of stimulus generalization. A detailed criticism of experiments of this general nature has been provided by Brown, Clarke, and Stein (1958).

This study attempted to extend Rosenbaum's findings by using a procedure that minimized the difficulties inherent in the earlier work: (a) the training procedure consisted of conditioning a single stimulus (an 8.9-cm. line); (b) *S* was given ample time to make a response; (c) the test series contained generalized stimuli (shorter lines), as well as several presentations of the training stimulus, but no booster trials. College students with high, middle, and low *MA* scale scores were subjected to one of three experimental conditions that closely resembled those used by Rosenbaum (1953), and it was hypothesized that increases in *MA* scale score and experimental anxiety would lead to greater generalization responsiveness.

METHOD

Subjects

Ninety men and 90 women were divided into thirds for each sex on the basis of the *MA* scale scores of an introductory psychology pool of 453 *Ss*. Ten of the 30 *Ss* in these six groups were randomly assigned to each of the three experimental anxiety conditions.

Stimuli

Seven lines, 5.0, 5.5, 6.1, 6.7, 7.4, 8.1, and 8.9 cm. long, were selected so that each successive length represented an approximate 10% increase over the length of the preceding line. Each of these stimuli was presented on a standard memory drum for 4 sec., followed by a 4-sec. blank interval. After a stimulus had been presented, *S* pulled a large handle with the right hand, and the experimenter (*E*) immediately recorded the amplitude of each response in pounds of pull.

Procedure

The *Ss* in the shock groups received stimulation through a clip placed on the fourth finger of the left hand. For *Ss* in the strong shock groups the current varied between 1.0 and 2.0 ma. ($M = 1.6$), and for *Ss* in the weak shock groups the current varied between .5 and .9 ma. ($M = .7$). The *Ss* in the buzzer groups were told that they would occasionally hear a buzzer. During a pretraining session *Ss* were trained to differentiate between a hard and a soft pull on the handle. For men a hard pull was defined as 35 lb. or more and a soft pull as less than 35 lb., and for women a hard pull was defined as 25 lb. or more and a soft pull as less than 25 lb. The *Ss* next were

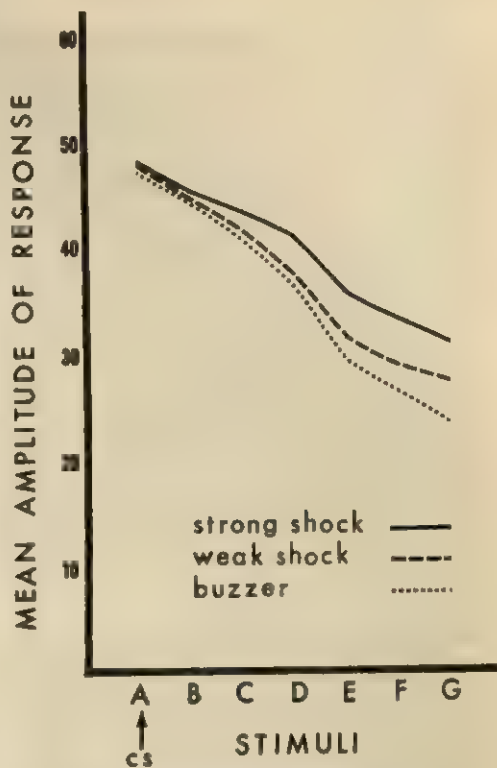


FIG. 1. Response amplitude for men differing in experimental anxiety.

given a sample of the shock or buzzer and told that their responding would be unrelated to its occurrence. Twenty training trials with the 8.9-cm. line ensued; if *S* made a hard pull he was told "right," and if he made a soft pull he was told "wrong." Stimulation occurred twice during these trials, randomly during Trials 1-5 and again during 11-15. The criterion for successful training was 8 of the last 10 responses correct (hard pulls).

Following training, the generalization stimuli, consisting of six randomized orders of the seven stimuli, were presented. Stimulation also occurred twice during these trials, randomly during Trials 8-15 and 21-28.

RESULTS

MA Scale Score

There were no systematic differences in either the heights or the slopes of generalization gradients as a function of *MA* scale score for men or for women.

Experimental Anxiety

Men's results. The gradients for men based on mean pounds of pull are presented in Figure 1. As the generalization stimuli become

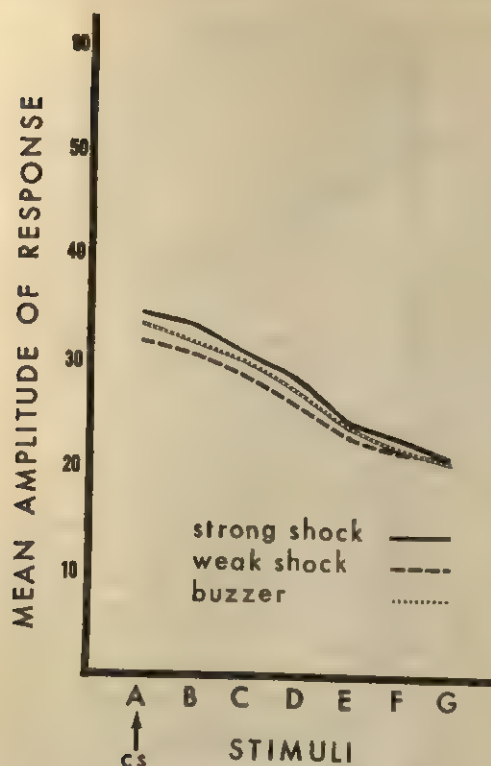


FIG. 2. Response amplitude for women differing in experimental anxiety.

more distant (shorter) from the training stimulus (A) all three curves decrease regularly. The gradient for the strong shock group is both highest and flattest, the gradient for the weak shock group is intermediate, and the gradient for the buzzer group is both lowest and steepest. Analysis of variance of these data reveals that, although there are no differences in height ($F = 1.7$, $df = 2/81$), differences in slope are highly significant ($F = 2.9$, $df = 12/486$, $p < .005$).

Women's results. The gradients for women are presented in Figure 2 and suggest smaller differences than were true for the men. Analysis of variance of these data reveals differences neither in the height ($F = 0.8$, $df = 2/81$) nor the slope ($F = 0.5$, $df = 12/486$) of the gradients.

Replications

Because there was a significant effect of experimental anxiety on the slope of generalization gradients for men but not for women, it was decided to attempt to replicate these

findings for the two most disparate male groups, strong shock and buzzer. The replication consisted of two identical phases, the first using the same *E* and the second a naive *E*. Each phase employed 20 additional male Ss from the same introductory psychology pool who were randomly assigned to the two conditions. Exactly the same experimental procedure as for the original male groups was followed.

Phase I. The gradients for the two groups in the first phase, presented in Figure 3, are similar in height and slope. Analysis of variance of these data indicate no significant differences.

Phase II. The gradients for the two groups run with the naive *E* in the second phase are presented in Figure 4. The gradient for the buzzer group is higher, suggesting that this group made *harder* pulls than the strong shock group. Slope differences appear to be minimal. The analysis of variance performed on these data again failed to achieve significance for either height ($F = 1.3$, $df = 1/18$) or slope ($F = 1.0$, $df = 6/108$). When the

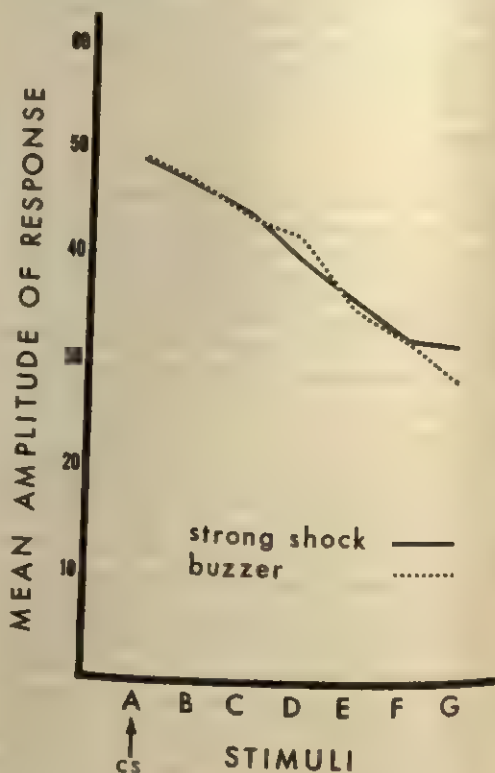


FIG. 3. Response amplitude for Replication Phase I.

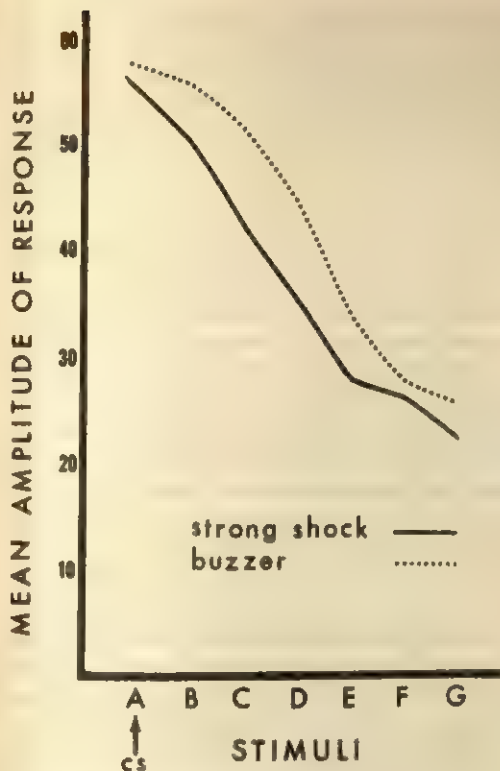


FIG. 4. Response amplitude for Replication Phase II.

data from two replications were pooled, the F values for both height and slope were less than 1. Thus there were no differences between the gradients, and the original slope differences for men failed to replicate.

DISCUSSION

MA Scale Score and Clinical Anxiety

The lack of a systematic relationship between MA scale score and stimulus generalization found in this study is consistent with the majority of previous research on the topic. In view of the fact that six other experiments (Arnhoff, 1959; Burns, 1964; Buss, 1955; Fager & Knopf, 1958; Mednick, 1957; Rothman, 1962) have also failed to demonstrate a relationship between either MA scale score or clinical anxiety and stimulus generalization it seems clear that evidence in support of such an hypothesis is lacking.

Experimental Anxiety

Although originally there was a flatter generalization gradient for the male strong shock

group than for the male buzzer group, this effect failed to replicate and proved to be unreliable. In only one other related study (Rosenbaum, 1953) has there been an attempt to vary anxiety experimentally, and in it a higher and flatter gradient was also obtained from the strong shock group. However, it has been pointed out that there were procedural difficulties in Rosenbaum's work, and, furthermore, the results of the present study are more general. The negative findings are not limited to situations demanding rapid responding, booster trials, or prior training with the generalization stimuli, and they apply to women as well as to men. These results, then, fail to substantiate the hypothesis that experimental or situational anxiety affects stimulus generalization gradients in a systematic manner.

REFERENCES

- ARNHOFF, F. N. Stimulus generalization and anxiety. *Journal of General Psychology*, 1959, 60, 131-136.
- BROWN, J. S., CLARKE, F. R., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *Journal of Experimental Psychology*, 1958, 55, 359-362.
- BURNS, J. T. Manifest anxiety and stimulus generalization under two sets of instructions. Unpublished master's thesis, Kent State University, 1964.
- BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *Journal of Abnormal and Social Psychology*, 1955, 50, 271-273.
- FAGER, R. E., & KNOPF, I. J. Relationship of manifest anxiety to stimulus generalization. *Journal of Abnormal and Social Psychology*, 1958, 57, 125-126.
- MEDNICK, S. A. Generalization as a function of manifest anxiety and adaptation to psychological experiments. *Journal of Consulting Psychology*, 1957, 21, 491-494.
- MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, 55, 316-327.
- ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *Journal of Experimental Psychology*, 1953, 45, 35-43.
- ROSENBAUM, G. Stimulus generalization as a function of clinical anxiety. *Journal of Abnormal and Social Psychology*, 1956, 53, 281-285.
- ROTHMAN, G. C. Stimulus generalization and discrimination following voluntary response training at various levels of anxiety. Unpublished master's thesis, Kent State University, 1962.
- TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, 48, 285-290.

(Received July 29, 1968)

EFFECTS OF TASK COMPLEXITY AND VERBAL EVALUATION ON THE LEARNING OF NORMALS AND SCHIZOPHRENICS¹

DAVID L. STREINER²

Syracuse University

This study tested the relative predictive powers of three theories of schizophrenia: Mednick's drive theory; Rodnick and Garmezy's censure-deficit theory; and Atkinson and Robinson's censure-sensitivity theory. Thirty normals, 30 process schizophrenics, and 30 reactive schizophrenics were each run in a low- and high-complexity verbal discrimination task. Half of each group was praised for correct responses, the other half was censured for incorrect ones. While normals did not respond differentially to the reinforcers, schizophrenics tended to learn faster when censured than when praised, supporting Atkinson and Robinson. The performance decrement from the low- to the high-complexity task was equivalent for all groups, which is nonsupportive of Mednick's theory. An alternative explanation of the results is offered and a cautious interpretation of the data is urged due to medication differences between normals and schizophrenics.

Within recent years, there have been two major classes of theories that attempt to explain differences in performance in learning tasks between normals and schizophrenics: drive theories and censure-sensitivity theories.

The major drive theory accepted by Western psychologists is that of Mednick (1958), which assumes that schizophrenics are in a state of drive that is higher than that of normals. Although Mednick's theory has come under serious criticism on both theoretical and empirical grounds (e.g., Buss, 1966), it has found wide acceptance in the United States, presumably due in large measure to its congruence with the familiar Hull-Spence theory.

Based on Hull's assumption of multiplicative relationship between drive (D) and habit strength (gH_R), Mednick predicts that in

simple tasks, schizophrenics, because of their higher drive, should perform better than normals. However, since D increases the strength of all responses, schizophrenics, in comparison with normals, should reach criterion more slowly and make a greater number of errors in complex tasks, which allow for a greater number of competing or alternative responses being brought above threshold and interfering with the desired response.

The censure-sensitivity theories of schizophrenia assume that the preschizophrenic person is born with or develops a heightened sensitivity to criticism, which has greater effects on his performance than on the performance of normals. One version of this theory, Rodnick and Garmezy's censure-deficit position (Garmezy & Rodnick, 1959; Rodnick & Garmezy, 1957), states that sensitivity to social censure leads to easy disruption of the schizophrenic's performance. Although much of the earlier research appeared to support their contentions, later studies have cast doubt upon the theory's validity (Cicchetti, Klein, Fontana, & Spohn, 1967; Fontana, Klein, & Cicchetti, 1967; Klein, Cicchetti, & Spohn, 1967).

Based on the same assumption of a heightened sensitivity to criticism in schizophrenia Atkinson and Robinson (1961) predict a dif-

¹ This article is based on a doctoral dissertation submitted to the Department of Psychology, Syracuse University. The author wishes to express his deep gratitude to the members of his committee, Sanford J. Dean (chairman), Wallace R. McAllister, and Norman Simonson. Special thanks are also due to Edward Parkes, Director of Psychological Services, and the staff of Norristown State Hospital, Norristown, Pennsylvania, whose cooperation made this study possible.

² Requests for reprints should be sent to David L. Streiner, Department of Psychiatry, Hamilton Psychiatric Hospital, P.O. Box 585, Hamilton, Ontario, Canada.

ferent effect of censure. They write,

For schizophrenic subjects, unlike normal subjects, the motivation to avoid mild punishment is stronger than the motivation to achieve mild reward, so that in situations in which giving a correct response is the only way to avoid punishment, punishment has a facilitating effect on performance [p. 325].

They report the results of two paired-associate learning experiments where schizophrenics improved with social censure.

Much of the variability in results can be explained by the vast difference in patient populations, in that the schizophrenics in one sample are rarely comparable to the schizophrenics in another sample. Despite the crudity of the Kraepelinian nosological system, even these distinctions in diagnoses are often ignored. Within recent years many investigators, if not diagnosticians, have abandoned the symptomatologically based Kraepelinian typology for a dichotomous "process-reactive" differentiation of schizophrenia based on the clinical history and probable prognosis of the patient.

Process schizophrenia involves a long-term progressive deterioration of the adjustment pattern with little chance of recovery, while reactive schizophrenia indicates a good prognosis based on a history of generally adequate social development with notable stress precipitating the psychosis [Herron, 1962, p. 329].

When schizophrenic Ss are divided into process (or "poor premorbid") and reactive (or "good premorbid") groups, the findings appear to be somewhat more consistent. Both Bleke (1955) and True (1966) reported results that show that normals and reactive schizophrenics are relatively insensitive to social censure, but that both groups differ significantly from the process schizophrenic group that did more poorly with censure, in support of the Rodnick and Garmezy position.

The purpose of the present study was twofold: to examine some of the variables that may account for the discrepancies in the findings, such as task complexity and schizophrenic diagnosis; and to establish conditions under which the drive and censure-sensitivity theories would predict divergent results in order to test the relative power of each of these theories.

METHOD

Design. The basic design for this study was a 3 (Diagnosis) $\times 2$ (Reinforcement) factorial, generating six experimental groups: normals receiving praise (N-P), normals receiving social censure (N-C), process schizophrenics receiving praise (P-P) and censure (P-C), and reactive schizophrenics receiving praise (R-P) and censure (R-C). Each S in each of the cells had two learning tasks, one of low complexity (LC) and one of high complexity (HC).

Subjects. In order to ascertain whether the HC task was in fact harder to learn than the LC task, 20 white male neuropsychiatric patients hospitalized at Norristown State Hospital, Norristown, Pennsylvania, were run in a pilot study, 10 patients per task.

The schizophrenic sample for the major study consisted of 60 white male neuropsychiatric patients, each with a primary diagnosis of schizophrenia. All of the patients were between the ages of 18 and 60 at the time of testing, had had at least a sixth-grade education, had no history of mental retardation or central nervous system pathology, were not currently alcoholic or addicted to drugs, and had not received insulin- or electro-shock therapy for at least 12 mo. prior to testing. The patients were divided into process and reactive groups on the basis of the Ullmann and Giovannoni (1964) Self-Report (S-R) Scale. All patients who scored between 0 and 12 were classified as process, and between 13 and 24 as reactive.

The process and reactive groups did not differ significantly on mean age at first hospitalization. At the time of testing, however, the process group had spent an average of 11.40 yr. in mental hospitals (not necessarily consecutive) and the reactive group had been institutionalized for an average of 5.47 yr. This difference was statistically significant at the .003 level ($F = 9.879$, $df = 1/56$).

The normal sample consisted of 30 white males between the ages of 18 and 60. All normal Ss had had at least a sixth-grade education, had no history of psychiatric hospital admissions, and were employees of Norristown State Hospital. There were no differences in the amount of education among the various diagnostic groups, but there was a significant age difference among the groups ($F = 24.998$, $df = 2/84$, $p < .0001$). A Scheffé test (Edwards, 1960) showed that the normals were significantly younger than the two schizophrenic groups, but that the process and reactive schizophrenic groups did not differ.

Stimuli. A verbal discrimination paradigm was used, in which a stimulus word and four response words were presented to Ss. In each case, two of the response words were "correct" and two "incorrect," although S's task was to learn any one of the correct responses. The stimulus word and the response words were taken from Palermo and Jenkins' (1964) norms, and all had A or AA ratings in Thorndike and Lorge's (1944) frequency count. In the LC task, the correct words had higher associative strengths to the stimulus word than did the incorrect words, while

CARS

BUSSES

DRIVE

FAST

TRUCKS

FIG. 1. Sample stimulus-response set shown to *S*.

the low associative strength words were correct in the HC task. That is, task complexity in this experiment was a function of the number of competing responses. In the LC task, the most dominant responses were correct, while in the HC task the most dominant responses were incorrect and competed strongly with the less dominant, correct responses that *S* had to learn.

For each stimulus-response set, the stimulus word was printed in $\frac{1}{2}$ -in. capital letters on the left side of a sheet of $8\frac{1}{2} \times 11$ in. white paper held horizontally, and the four response words were printed in a columnar fashion on the right side. The ordering of the four words was randomized for each of the sets. Thirty-five millimeter black and white slides were taken of each set. The 10 LC sets were randomized seven times and placed in a Kodak Carousel slide holder with a completely opaque slide separating each trial of 10 stimulus-response sets. Thus, the order of sets on the eighth and fifteenth trials was identical to the order on the first trial, the ninth and sixteenth orders were identical with the second, and so on. Separate carousels were made for the 10 LC and 10 HC sets.

Procedure. For the pilot study, each *S* received only one of the tasks, and all *Ss* were tested under the praise condition. The *Ss* were alternately assigned to the LC and HC conditions in order of their appearance.

All *Ss* in the major study were seen individually in a private office. At the beginning of the first testing session, each schizophrenic *S* was given the process-reactive S-R Scale. Since many of the patients misinterpreted some of the items when the scale was self-administered, the items were read aloud to all *Ss* by the experimenter (*E*). For each diagnostic category, *Ss* were alternately assigned to the praise or censure condition in order of their appearance. In the praise condition, correct responses were reinforced by *E* saying "right" and incorrect choices by *E* not saying anything. In the censure condition, *E* said "wrong" after incorrect responses and nothing for correct ones. All *Ss* were then given the following instruction (modifications of the instructions for the censure-condition *Ss* are given in parentheses):

I'm going to show you a series of slides on the wall. Each slide will look like this. (*S* was shown a 3×5

index card as in Figure 1.) There will be a single word on the left side and four words on the right side. Your job will be to learn which *one* of the four words goes with the word on the left. Here's how we'll do it. As each slide is presented, say the word on the left aloud, in this case, "cars," and then tell me which one of the four words you think is right. If you've chosen the correct word, let's say "drive," I will say "right" (I won't say anything). If you *haven't* chosen the correct word, I won't say anything (I will say "no" or "wrong"). If you chose the wrong word, don't try another word just yet. There will be ten such sets of words, then a brief pause, and then we'll go through the same set of words, but in a different order. If you chose the right word the first time, stick with it; that is, say "cars-drive" again, but if you didn't get the word right the first time, try a different word. We'll go through the list a number of times until you've gotten all the words correct three times in a row. Any questions?

Because it was expected that all groups would learn the LC task faster than the HC task, the order of task presentation remained constant for all *Ss*. In this experiment, counterbalancing would most probably have led to greater intragroup variability, in that half of the *Ss* would start with a task that they solve relatively rapidly and half with a task that takes longer to learn. Consequently, some *Ss* would begin the second task with a history of relatively rapid success and some with a history of relatively slower learning and possibly more frustration. Furthermore, a fixed order of presentation would favor an equivalent amount of transfer from Task 1 to Task 2 for all *Ss*.

For the LC task, the internal timer in the projector was set for a 15-sec. display time. The HC task was given at least 24 hr. after the LC task to minimize fatigue factors. The response contingencies remained the same for each *S*, but the projector was set for an 8-sec. exposure. Each *S* was reminded of the task and of the response contingencies.

For both tasks, *Ss* were run to a criterion of three consecutive errorless trials. All *Ss* who did not reach criterion within 30 trials were eliminated from the study.

RESULTS

To determine if the HC task was in fact harder than the LC task, an analysis of variance was performed on the number of correct responses on the first trial of the major study for both tasks. The analysis showed that significantly more correct words were chosen as first responses on the LC task than on the HC task ($F = 37.44$, $df = 1/84$, $p < .0001$).³

³ The pilot study showed it took twice as many trials to learn the HC task ($\bar{X} = 15.6$) than the LC task ($\bar{X} = 7.1$) ($t = 3.29$, $df = 18$, $p < .001$).

For the major study, a Lindquist Type III analysis of variance (Lindquist, 1953) was performed on the number of trials to criterion, in which task was the repeated measure. The analysis yielded significant main effects of diagnosis ($F = 11.002$, $df = 1/84$, $p = .0002$), reinforcement ($F = 4.842$, $df = 2/84$, $p = .029$), and task ($F = 16.317$, $df = 1/84$, $p = .0003$). A Scheffé test revealed that the significant diagnosis effect was caused by the superiority in learning speed of the normal group, which differed significantly from the two schizophrenic groups, while the two schizophrenic groups did not differ from each other.

The significant reinforcement effect showed that fewer trials were needed to reach criterion under the censure condition ($\bar{X} = 8.58$) than under the praise condition ($\bar{X} = 9.99$), while the task effect reflected fewer trials necessary to learn the LC task ($\bar{X} = 8.47$) than the HC task ($\bar{X} = 10.10$). The Task \times Reinforcement interaction approached significance ($F = 2.997$, $df = 1/84$, $p = .083$), reflecting a tendency for Ss in the praise condition to show more of a performance decrement from the LC to the HC task than Ss in the censure condition. Analysis of each task individually showed that there was no effect in the HC task. Thus, on relatively simple tasks, the speed of learning was independent of the type of reinforcement, but, under more difficult conditions, Ss learned faster with censure than with praise.

Although the main analysis showed that the Reinforcement \times Diagnosis interaction did not approach an accepted level of significance ($F = 1.666$, $df = 2/84$, $p = .193$), examination of Figure 2 indicates that there is a tendency for a differential effect of reinforcement on diagnostic category. Combining both schizophrenic groups yielded a stronger interaction ($F = 2.90$, $df = 1/86$, $p = .09$). Thus, while normals perform equally well under conditions of praise and censure, there is a tendency for schizophrenics to learn better when censured for wrong responses than when praised for correct ones.

The Task \times Diagnosis interaction, predicted by Mednick's theory, was neither appreciable nor statistically significant ($F =$

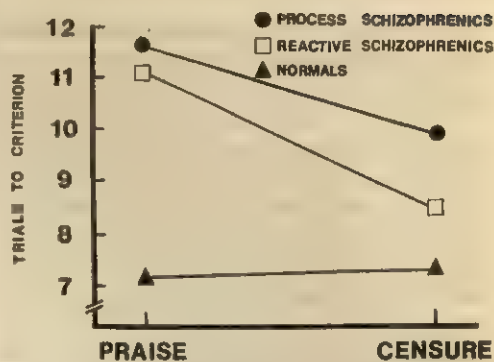


FIG. 2. Number of trials to criterion as a function of reinforcement for the three diagnostic groups.

0.007 , $df = 2/84$, $p = .994$), indicating that the performance decrement from the LC to the HC task was about equal for all diagnostic groups; that is, there was no differential effect of task complexity on the performance of the different groups.

Since it is possible that patients' responsiveness to social criticism becomes flattened and dulled after extended hospitalization, the analysis was repeated using long-term versus short-term hospitalization rather than the process-reactive dichotomy. For both the praise and censure groups, the patients were divided into groups by a median split ($Mdn = 5$ yr. of hospitalization), irrespective of the patients' process-reactive scores. In the short-term group, there were 19 reactive and 11 process schizophrenics, and 19 process and 11 reactive schizophrenics in the long-term group. The results of this analysis were directly comparable with those of the previous analysis, which used process-reactive as a dimension. Despite the fact that there was not a complete overlap between the reactive and short-term groups, or between the process and long-term groups, long-term schizophrenics performed in a manner similar to process schizophrenics, and short-term schizophrenics performed similarly to reactive schizophrenics.

DISCUSSION

The findings of this study lend partial support to the censure-sensitivity theories that predict better performance by schizophrenics under conditions of censure than of praise.

The results did not, however, support the drive interpretation of schizophrenic performance deficit.

While the interaction between reinforcement and diagnosis approached significance only when the schizophrenic groups were combined and compared with normals, this tendency was in the direction predicted by the Atkinson and Robinson version of censure-sensitivity theory. Moreover, there was a strong tendency for the differential advantage of censure to be more pronounced on harder tasks than on easier one. At the same time, normals did not respond differentially to praise or censure, regardless of the complexity of the task involved. These findings, therefore, are contrary to the interaction predicted by Garnezy and Rodnick, who state that schizophrenics' performance should deteriorate under conditions of response-contingent social censure.

Although the absence of "right" in the praise or R-N condition is functionally equivalent to and conveys the same message as saying "no" or "wrong" in the censure or N-W condition, it appears that, for schizophrenics, the actual word "wrong" or "no" has a greater psychological effect than the "wrong" that is implied by silence in the praise condition. Thus, while the normal person may respond largely to the information implied by the words, the schizophrenic may have been responding more to their evaluative connotations.

This finding, though, is open to two interpretations. The first, and that taken by Atkinson and Robinson, would state that the schizophrenics' performance improved with response-contingent censure, and that praise had no effect or may have had a relatively mild detrimental effect. This presupposes, however, that there is an initial deficit in the performance of schizophrenics that was partially overcome by their motivation to avoid mild censure. The alternative explanation, which would also be congruent with the results shown in Figure 2, is that there is no initial deficit in the schizophrenics' performance, and that their speed of learning is equivalent to that of normals. Praise, however, may not serve as an effective reinforcer for

schizophrenics, and they show slower learning under this condition.

If the performance of the normal group is taken as a base line, then the data reported here favor the second interpretation. However, if the schizophrenics' performance under conditions of praise is used as a basal level, then the data support the first interpretation. The difficulty resides in the choice of an adequate base with which schizophrenics' performance under different conditions may be compared.

The failure to support Mednick's theory raises the question of the adequacy of the study to test his hypothesis. Of all the differences between normals and schizophrenics, perhaps none is as crucial as the iatrogenic difference that, for the most part, schizophrenics receive psychoactive tranquilizers and mood elevators, and normals do not. Moreover, if the assumption is made that the dosage of tranquilizing agents is proportional to the schizophrenic's state of overt agitation and symptomatology, then the schizophrenics with the greatest amount of drive also receive the largest amounts of tranquilizers. The result would be to decrease the drive-level differences among schizophrenics and bring the drive level of schizophrenics down to a level more nearly approximating that of normals. Consequently, schizophrenics would perform in much the same way as normals, except that they would be more susceptible to associative interference due to the tranquilizing effects of the drugs. In effect, therefore, they would respond to LC and HC tasks as do normals, but show an overall decrease in performance, as was demonstrated here.

The lack of any difference in performance between process and reactive schizophrenics may also be a function of the medication. Most of the studies that have reported significant differences between process and reactive schizophrenics were carried out in the early or mid 1950s, whereas the major tranquilizers were introduced into hospitals in significant quantities in the late 1950s and early 1960s. Thus, while the two groups still differ in terms of their premorbid histories, the drugs may have artificially eliminated concomitant differences in cognitive functionings, such as

those due to flight of ideas. Unfortunately, due to the wide diversity of medication, it was impossible to evaluate the effect of the drugs on performance.

While the results of this study tend to support the contention that schizophrenics' reactions to praise and censure differ from normals' reaction, and do not support the drive theory of schizophrenia, a cautious interpretation of the data is necessary.

REFERENCES

- ATKINSON, R., & ROBINSON, N. M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, 62, 322-326.
- BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *Journal of Personality*, 1955, 23, 479-498.
- BUSS, A. H. *Psychopathology*. New York: Wiley, 1966.
- CICCHETTI, D. V., KLEIN, E. B., FONTANA, A. F., & SPOHN, H. E. A test of the censure-deficit model in schizophrenia, employing the Rodnick-Garmezy visual-discrimination task. *Journal of Abnormal Psychology*, 1967, 72, 326-334.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.
- FONTANA, A. F., KLEIN, E. B., & CICCHETTI, D. V. Censure sensitivity in schizophrenia. *Journal of Abnormal Psychology*, 1967, 72, 294-302.
- GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease*, 1959, 129, 450-466.
- HERRON, W. G. The process-reactive classification of schizophrenia. *Psychological Bulletin*, 1962, 59, 329-343.
- KLEIN, E. B., CICCHETTI, D., & SPOHN, H. A test of the censure deficit model and its relation to premorbidly in the performance of schizophrenics. *Journal of Abnormal Psychology*, 1967, 72, 174-181.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin*, 1958, 55, 316-327.
- PALERMO, D. S., & JENKINS, J. J. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. *Nebraska Symposium on Motivation*, 1957, 5, 109-184.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.
- TRUE, J. E. Learning of abstract responses by process and reactive schizophrenia patients. *Psychological Reports*, 1966, 18, 51-55.
- ULLMANN, L. P., & GIOVANNONI, J. M. The development of a self-report measure of the process-reactive continuum. *Journal of Nervous and Mental Disease*, 1964, 138, 38-42.

(Received July 29, 1968)

THE MMPI AND RECIDIVISM

JAMES L. MACK¹

University of Minnesota

MMPIs of 80 recidivists and 68 parole successes from a state training school were compared by analyzing mean scale scores, elevations, code types, and mean ranks of the clinical scales. Recidivists were expected to be characterized by elevations and code types reflecting the relative importance of the *F*, *Pd*, and *Ma* scales and the nonrecidivists by the *Hs*, *D*, *Hy*, and *Si* scales. No important differences were identified. Conclusions are: (a) the MMPI used alone is not useful in identifying recidivists in relatively homogeneous delinquent populations; (b) its use should be restricted to the exploration of small differences between groups; and (c) the combined use of historical information and the MMPI to identify recidivists needs study.

A number of investigators have attempted to distinguish recidivists from individuals who successfully completed parole from a correctional institution by means of the Minnesota Multiphasic Personality Inventory (MMPI). Negative results have been reported with single predictor scales (Freeman & Mason, 1952; Gough, Wenk, & Rozyner, 1965; Mandel & Barron, 1966; Watron, 1963); correlational studies (Clark, 1948); mean profile comparisons (Panton, 1962); blind sorting by clinical judges (Mandel & Barron, 1966); and comparison of proportions of abnormal profiles (Clark, 1948). Investigators using large *N*s have obtained significant, though small, group differences on the *F*, *Hs*, *D*, *Hy*, *Pd*, *Ma*, and *Si* scales (Gough et al., 1965; Wirt, 1967), leading Wirt to characterize recidivists as more sociable and psychopathic and less neurotic than parole successes. Most previous investigators used a single approach in analyzing MMPI data, and none attempted to evaluate configural aspects of individual profiles.

The present investigation represents an attempt to provide conclusive results concerning the effectiveness of the MMPI in distinguishing recidivists from nonrecidivists by compar-

ing (a) mean scale scores, (b) the frequency of abnormal profiles, (c) the proportions of various code types within each group, and (d) the mean ranks of each scale. In general it is predicted that the recidivists will be distinguished by the relative importance of the *F*, *Pd*, and *Ma* scales and the nonrecidivists by the *Hs*, *D*, *Hy*, and *Si* scales.

METHOD

Subjects. All boys from Hennepin and Ramsey Counties, Minnesota (the primary metropolitan area of the state), who were paroled from the State Training School at Red Wing between January 1, 1966, and July 31, 1966, were included in the study. Of the 159 *S*s, 82 were recidivists, 69 nonrecidivists, and 8 remained unclassified. There was no difference between the groups with respect to age ($M = 16.9$ yr., range = 14-19), IQ, and length of time institutionalized ($M = 5.2$ mo., range = 1-13). The majority of *S*s were of normal intelligence, only 2 falling below an IQ of 70. The nonrecidivists were on parole for an average of 9.1 mo. (range = 4-16), while the recidivists had been on parole for an average of 4.2 mo. (range = 1-12).

Procedure. Previous to his placement at the training school, each *S* was administered the MMPI as part of a routine battery of tests at a reception center. Testing was thus carried out more than 5 mo. prior to the parole of all *S*s and, in some cases, more than 12 mo. earlier. Because of the possible effects of parole or institutional adjustment on test performance, no attempt was made to obtain more recent MMPI results. The decision of whether or not *S* had failed on parole was made by the Youth Commission of the Minnesota Department of Corrections upon the recommendation of the agent supervising the *S* on parole. The status of eight *S*s could not be clearly ascertained, and they were left unclassified.

¹ This article is based on a doctoral dissertation submitted to the University of Minnesota, 1967, and was prepared with the aid of Robert Wirt, adviser, and James Butcher and Peter Briggs, who served on the examining committee.

Requests for reprints should be sent to James L. Mack, Norristown State Hospital, Norristown, Pennsylvania 19401.

RESULTS

There were no invalidating *K* or *L* scores (*T* scores > 75). High *F* scores in themselves were not considered invalidating in order to retain valid clinical information (Morrice, 1957). Three *Ss* produced profiles suggesting inability to understand test items and were omitted, leaving 80 recidivists and 68 non-recidivists.

When the groups were compared on the basis of their mean *T* scores on the 3 validity and 10 clinical scales, there were no significant differences at even the .10 level. These data are presented in Table 1. At the .10 level of significance the probability that the research design would detect a difference as small as 5 *T* scores was .94. For a difference of 2.5 *T* scores, however, the probability of detection was only .42, indicating that the sample size was not sufficient to allow the identification of small group differences with any degree of confidence.

All profiles were then coded according to the Welsh system (Welsh, 1948). The proportions of individual code types within each of the groups were compared in regard to two point codes exclusive of elevation, codes of all *T* scores greater than 69, and total codes. All codes were heavily weighted with elevations on the *Pd* scale. The recidivists had the

TABLE 1
MMPI MEAN *T* SCORES

Scale	Group				Z
	Nonrecidivists		Recidivists ^a		
	M	SD	M	SD	
L	48.60	6.80	49.11	7.31	.44
F	61.12	8.42	61.06	9.24	.04
K	52.07	8.56	53.16	8.18	.78
Hs	55.53	10.73	55.08	9.16	.27
D	63.12	12.99	60.90	11.93	1.08
Hy ^b	59.32	8.97	57.28	7.27	1.50
Pd	75.44	8.65	74.48	9.05	.66
Mf	53.37	8.81	54.05	8.80	.47
Pa	62.35	9.47	61.43	9.27	.59
Pt	66.97	10.20	65.10	10.43	1.10
Sc	66.94	11.76	68.24	13.34	.63
Ma	65.46	11.70	63.50	11.31	1.03
Si	53.40	10.05	52.39	8.78	.64

^a *N* for the *Pa* scale was 79.

^b Variances for the *Hy* scale were significantly different ($p < .10$).

TABLE 2
MEAN RANKS OF MMPI SCALES

Scale	Group				Z
	Nonrecidivists ^a		Recidivists ^b		
	M	SD	M	SD	
<i>Hs</i>	7.4	2.15	7.3	2.10	.28
<i>D</i>	5.2	2.86	5.5	2.71	.65
<i>Hy</i>	6.2	2.13	6.3	2.30	.27
<i>Pd</i>	2.2	1.58	2.3	1.71	.36
<i>Mf</i>	7.7	2.00	7.4	2.40	.83
<i>Pa</i>	5.4	2.21	5.3	2.28	.27
<i>Pt</i>	4.3	2.19	4.4	2.12	.28
<i>Sc</i>	4.3	2.29	3.7	2.40	1.54*
<i>Ma</i>	4.6	2.98	4.9	2.89	.61
<i>Si</i>	7.6	2.55	7.8	2.22	.50

^a *N* = 68.

^b *N* = 79.

* $p < .13$.

Pd scale as a high point in 41 of 80 profiles and the nonrecidivists in 33 of 68 profiles. The groups were quite similar in almost all comparisons. The recidivists did show a greater proportion of profiles with *Pd* and *nothing else* greater than 69 ($p = .036$). The nonrecidivists had a greater proportion of elevations greater than 69 on the *Hy* scale ($p = .036$).

To further evaluate profile configuration the mean rank of each scale was calculated for each group and comparisons made. No significant differences were observed, although the recidivists showed a slight tendency to have the *Sc* scale more highly ranked within their profiles ($p < .13$). Again, the predominant influence of the *Pd* scale was evident.

DISCUSSION

Previous studies have been criticized for failing to identify all recidivists. All non-recidivists in the present study had been on parole for longer than 6 mo., and statistics of the Minnesota Department of Corrections² indicate that 80% of those who eventually violate their parole do so within the first 6 mo. of parole. Furthermore, the proportion

² W. V. Sittler, R. L. Webb, and N. G. Mandel. "A follow-up study of parole violation of individuals released from Minnesota correctional facilities from January 1, 1964 through December 31, 1964." Unpublished manuscript, Minnesota Department of Corrections, St. Paul, 1966.

of parole violators in the present sample, 49.1%, compares to an overall rate of 48.7% from the previous year's releases from the training school, lending weight to the assumption that Ss were correctly classified.

Although the *N* was not sufficiently large to reveal small group differences, attempting to enlarge the sample by taking parolees released over a longer time period, from a wider geographical area, or from other institutions would have introduced unwanted sources of variance. The results of the present investigation strongly suggest that the MMPI, considered alone, is not associated with recidivism to any important extent within such homogeneous populations as the present one, although the existence of a minimal association, of doubtful predictive utility, cannot be ruled out. That two comparisons of more than 50 showed differences significant at the .05 level can scarcely be considered telling evidence for a relationship between the MMPI and recidivism. Both groups in the present study showed a considerable degree of personality disturbance, but they did not appear particularly different with regard to their personality adjustment. Further research with the MMPI and recidivism might well be restricted to samples of very large size where the purpose is to demonstrate very small group differences.

Perhaps the most serious criticism that can be leveled at the present study is that no attempt was made to control for the effects of the environmental conditions into which the parolee was placed upon release. Nearly all were returned to their parental homes, the character of which may have been instrumental in influencing the parole adjustment of the Ss. The evaluation of MMPI results in combination with historical information con-

cerning the character of the parental home has been carried out in several studies of delinquency (Briggs, Wirt, & Johnson, 1961; Wirt & Briggs, 1959). Applications of these techniques may prove to be the most fruitful approach in future studies of recidivism.

REFERENCES

- BRIGGS, P. F., WIRT, R. D., & JOHNSON, R. An application of prediction tables to the study of delinquency. *Journal of Consulting Psychology*, 1961, 25, 46-50.
- CLARK, J. H. Application of the MMPI in differentiating A.W.O.L. recidivists from nonrecidivists. *Journal of Psychology*, 1948, 26, 229-234.
- FREEMAN, R. M., & MASON, H. M. Construction of a key to determine recidivists from nonrecidivists using the MMPI. *Journal of Clinical Psychology*, 1952, 8, 207-208.
- GOUGE, H. G., WENK, E. A., & ROZYNKO, V. V. Parole outcome as predicted from the CPI, the MMPI, and a base expectancy table. *Journal of Abnormal Psychology*, 1965, 70, 432-441.
- MANDEL, N. G., & BARRON, A. J. The MMPI and criminal recidivism. *Journal of Criminal Law, Police Science, and Criminology*, 1966, 57, 35-38.
- MORRICE, J. K. W. The MMPI in recidivist prisoners. *Journal of Mental Science*, 1957, 103, 632-635.
- PANTON, J. H. The identification of habitual criminalism with the MMPI. *Journal of Clinical Psychology*, 1962, 18, 133-136.
- WATTRON, J. B. A prison maladjustment scale for the MMPI. *Journal of Clinical Psychology*, 1963, 19, 109-110.
- WELSH, G. S. An extension of Hathaway's MMPI profile coding system. *Journal of Consulting Psychology*, 1948, 12, 343-344.
- WIRT, R. D. MMPI characteristics associated with outcomes in group psychotherapy with prisoners. Paper presented at the Second Conference on Recent Developments in the Use of the MMPI, Minneapolis, Minnesota, April 1967.
- WIRT, R. D., & BRIGGS, P. F. Personality and environmental factors in the development of delinquency. *Psychological Monographs*, 1959, 73(15, Whole No. 485).

(Received August 2, 1968)

RELATIONSHIP BETWEEN PERCEPTUAL DEPRIVATION TOLERANCE AND ADEQUACY OF DEFENSES AS MEASURED BY THE RORSCHACH¹

NANCY A. WRIGHT AND JOHN P. ZUBEK²

University of Manitoba

The Buhler-Lefever standardization of the Rorschach was administered, pre-experimentally, to 18 Ss who successfully endured 1 wk. of perceptual deprivation and to 13 Ss, the quitters, who were unable to do so. An Index of Control based on the Effectiveness of Defense in controlling Defense Demand was derived for each S. The results revealed that the group with a high-control Index was composed largely of successful Ss whereas the group with a low-control Index consisted largely of isolation quitters. It was concluded that the Index of Control, derived from the Rorschach, may serve as an excellent predictor of isolation tolerance.

It has long been known that not all volunteers can endure prolonged periods of perceptual deprivation in an isolation chamber. Some can tolerate it for many days while others terminate the condition after only a few hours (Zubek, 1964). Despite the large individual differences, no satisfactory predictive measures of deprivation tolerance have as yet been developed (see reviews of Myers, 1969; Schultz, 1965). Various paper-and-pencil tests of personality (e.g., Minnesota Multiphasic Personality Inventory, Edwards Personal Preference Schedule, and Thurstone's Temperament Schedule), together with certain perceptual measures, have been employed, but little success has been achieved in differentiating between Ss who can and cannot tolerate prolonged isolation. Furthermore, any significant results that have emerged have frequently contradicted one another.

Although paper-and-pencil tests of personality have proved to be largely ineffective, Wright and Abbey (1965), in an exploratory study conducted at the Manitoba Isolation Laboratory, have suggested that the prediction of perceptual deprivation tolerance could be improved greatly by examining the Rorschach manifestations of an individual's defense and control mechanisms and subse-

quently deriving his Index of Control. This Index appeared to be a reliable and valid means of determining whether or not drive-dominated material emerging during isolation could be controlled and, hence, whether or not S could endure isolation for the prescribed period of 1 wk.

There was some support for this idea in the literature at that time (Christy & Rasmussen, 1963; Nardini, Herrman, & Rasmussen, 1962). In a more recent article (Kracke, 1967), Erika Fromm is quoted as stating that the isolation situation created in some of the experiments is one that is as close as an experimental situation can get to the intra-uterine situation. Why, she asks, is it not sheer bliss? Kracke concludes that for many Ss the resulting stress is due to the ascendancy of primary process material in Ss for whom this is ego threatening.

The original findings of Wright and Abbey (1965) were considered to be tentative since the 21 male Ss were tested several months *after* completion of perceptual deprivation and thus their responses may have been affected by their feelings of success or failure in the situation. It was therefore decided to replicate the study in the same laboratory, under the same conditions of perceptual deprivation, employing the same duration (1 wk.), but instituting *prior* testing procedures. It was again hypothesized that the Rorschach records of successful Ss (those completing 1 wk. of isolation) would be characterized by

¹ This research was supported by the Defense Research Board, Canada (No. 9425-08) and by the National Research Council, Canada (APA-290).

² Requests for reprints should be sent to John P. Zubek, Department of Psychology, University of Manitoba, Winnipeg, Canada.

TABLE 1

DIVISION OF SUCCESSFUL AND UNSUCCESSFUL ISOLATION SUBJECTS ON THE BASIS OF CONTROL OF DRIVE-DOMINATED RESPONSES DERIVED FROM THE RORSCHACH

Type of Ss	Index of Control		
	High	Low	Total
Successful	14	4	18
Unsuccessful (quitters)	1	12	13
Total	15	16	31

drive-dominated material that was not threatening and was well controlled and that the Rorschach records of unsuccessful Ss, or quitters, would be characterized by lack of control so that the drive-dominated material would become conscious and cause mobilization of anxiety and maladaptive and pathological defenses.

METHOD

A group of 31 male university students were placed individually in an air-conditioned isolation chamber for a prescribed period of 1 wk. Of this group, 18 successfully endured the week, while 13, the unsuccessful Ss or quitters, were unable to do so (M duration = 56 hr.). All Ss wore translucent goggles, which permitted diffuse light but eliminated pattern vision, and a set of earmuffs through which white noise was constantly presented somewhat above the threshold of hearing. Further details of the isolation procedure are available in an earlier publication (Zubek, Aftanas, Hasek, Sansom, Schludermann, Wilgosh, & Winocur, 1962).

As part of the routine personality testing that preceded isolation, the senior author administered the Buhler-Lefever (Buhler, Buhler, & Lefever, 1954) standardization of the Rorschach. This procedure directs S to give from three to five responses per card and it encourages him to elaborate in the inquiry. The number of responses and the amount of material are therefore roughly equivalent from S to S . The Rorschach protocols were then analyzed 2 yr. after their preexperimental administration, a procedure that ensured that the tester had no knowledge of which S endured or failed to endure the week of perceptual deprivation. As in the earlier exploratory study, the protocols were scored objectively according to the Holt-Havel (Holt & Havel, 1961)³ method for assessing primary and secondary process in the

Rorschach and the defense and control mechanisms that these demand. The proportion of the Effectiveness of Defense (DE) score to the Defense Demand (DD) score became the Index of Control (Wright & Abbey, 1965).

RESULTS

The 31 Ss were rank ordered according to their Index of Control. They were then split at the median, into a high-control group (Index range = .47-.70; $n = 15$) and a low-control group (Index range = .78-1.50; $n = 16$). Table 1 shows the number of successful and unsuccessful Ss in each group. It can be seen that in the high-control group, 14 of the 15 Ss successfully completed the 1-wk. isolation period whereas in the low-control group, only 4 of the 16 Ss were able to do so. Since the expected frequency was $n < 5$ and $n < 20$, the median test for χ^2 was used (Siegel, 1956). This analysis demonstrated that the effectiveness of S 's control and defense mechanisms was positively related to success or failure in isolation ($\chi^2 = 12.17$, $df = 1$, $p < .01$).

DISCUSSION

Since the Index of Control in both this study and in the earlier exploratory report (Wright & Abbey, 1965) has clearly differentiated between the two types of isolation Ss, it appears to be a useful measure for predicting whether or not an individual's defense and control mechanisms will be effective in controlling drive-dominated responses that might emerge during his attempt to successfully complete a prolonged period of perceptual deprivation. Thus, at long last a satisfactory method of predicting isolation tolerance, by means of prior psychological testing procedures, seems to have been developed. This is a long-awaited finding and it is hoped that this method will be used by investigators in other laboratories in an attempt to further replicate these findings and develop normative data. Furthermore, this method also may prove useful in research on other types of stressful conditions in which the objective is to differentiate between Ss who may or may not be able to endure them.

It is interesting that in the present group of 31 Ss, the Index of Control scores tended to be better, that is, higher, than those in the

³ See also R. R. Holt, Manual for the scoring of primary process manifestations in Rorschach responses. Research Center for Mental Health, New York University. Unpublished draft of an experimental procedure, 1963.

original group of 21 Ss who were tested following their isolation experience. It is as if the defense and control mechanisms were on the alert immediately prior to entering such a stressful situation, and the efforts at control were therefore more successful. The scores of the low-control group in the original study also may have been depressed further by the feeling of failure that clouded the testing situation.

Finally, these findings have some implications for those investigators concerned with the beneficial or therapeutic effects of isolation. For example, Brownfield (1965), Shurley (1966), and Adams, Robertson, and Cooper (1966) have emphasized the importance of determining which personality characteristics are essential for a relaxed and pleasant adaptation to isolation. If, in fact, the control of emergent drive-dominated material is the crucial characteristic, which the present study suggests that it may be, such investigators should employ a prior Rorschach testing procedure using the Holt-Havel scoring method and the Wright-Abbey Index of Control.

REFERENCES

- ADAMS, H. B., ROBERTSON, M. H., & COOPER, D. Sensory deprivation and personality change. *Journal of Nervous and Mental Disease*, 1966, 143, 256-265.
- BROWNFIELD, C. A. *Isolation: Clinical and experimental approaches*. New York: Random House, 1965.

- BUHLER, C., BUHLER, K., & LEFEVER, D. W. *Rorschach standardization studies*. Los Angeles: Western Psychological Services, 1954.
- CHRISTY, R. L., & RASMUSSEN, J. E. Human reliability implications of the U.S. Navy's experience in screening and selection procedures. *American Journal of Psychiatry*, 1963, 120, 540-547.
- HOLT, R. R., & HAVEL, J. A method for assessing primary and secondary process in the Rorschach. In M. A. Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York: Wiley, 1961.
- KRACKE, W. H. The maintenance of the ego: Implications of sensory deprivation research for psychoanalytic ego psychology. *British Journal of Medical Psychology*, 1967, 40, 17-27.
- MYERS, T. I. Tolerance for sensory and perceptual deprivation. In J. P. Zubek (Ed.), *Sensory deprivation: Fifteen years of research*. New York: Appleton-Century-Crofts, 1969.
- NARDINI, J. E., HERRMAN, R. S., & RASMUSSEN, J. E. Navy psychiatric assessment program in the Antarctic. *American Journal of Psychiatry*, 1962, 119, 97-105.
- SCHULTZ, D. P. *Sensory restriction*. New York: Academic Press, 1965.
- SHURLEY, J. T. Stress and adaptation as related to sensory/perceptual isolation research. *Military Medicine*, 1966, 131, 254-258.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- WRIGHT, N. A., & ABBEY, D. S. Perceptual deprivation tolerance and adequacy of defenses. *Perceptual and Motor Skills*, 1965, 20, 35-38.
- ZUBEK, J. P. Effects of prolonged sensory and perceptual deprivation. *British Medical Bulletin*, 1964, 20, 38-42.
- ZUBEK, J. P., APTANAS, M., HASEK, J., SANSOM, W., SCHLUDERMANN, E., WILGOSH, L., & WINOCUR, G. Intellectual and perceptual changes during prolonged perceptual deprivation: Low illumination and noise level. *Perceptual and Motor Skills*, 1962, 15, 171-198.

(Received August 5, 1968)

THREE FACES OF EVE REVISITED: A STUDY OF TRANSIENT MICROSTRABISMUS

WILLIAM S. CONDON,¹ WILLIAM D. OGSTON, AND LARRY V. PACOE

Western Psychiatric Institute and Clinic, University of Pittsburgh Medical School

A film of a patient with multiple personality (the original Thigpen and Cleckley's patient, in a film called "Three Faces of Eve") was examined frame by frame. Each of the three personalities of the patient emerged in this film. A micro-transient strabismus was detected and described, and its occurrence in relation to each of the three personalities analyzed. Eve Black, the least stable personality, exhibited five times as many of these strabismic occurrences as the other two personalities. A somewhat similar strabismus has also been observed in some films of schizophrenic patients. This strabismus has not been observed thus far in films of normal behavior.

Thigpen and Cleckley (1954) first reported a case of multiple personality about which they later wrote a book entitled *The Three Faces of Eve*, which was produced in Hollywood, dramatizing the original case material. Thigpen and Cleckley themselves made a 30-min. sound film of the actual patient in her three different personalities and at stages of her therapy. This article is based on findings derived from this latter film,² using a method of film analysis described more fully elsewhere (Condon & Ogston, 1967).

In actuality, four personalities were reported by the authors, since one of the three, the final personality, Jane, displayed two aspects. The initial presenting personality was Eve White, a demure, shy, and well-mannered woman, although sadness and tension could be seen behind her reserve. Eve Black, the first of the multiple personalities to emerge, was a mischievous hoyden. Jane, who revealed two aspects (Jane I and Jane II), seemed to be a mature and capable person. She emerged late in therapy and ultimately became the dominant personality (Thigpen & Cleckley, 1957).

METHOD

The overall method of film analysis was essentially ethological in nature. The film was viewed over and over in the search for possible regularities. The

transformations of expressive behavior, especially facial, among the three personalities Eve White, Eve Black, and Jane, as described by Thigpen and Cleckley (1957), were clearly evident on a frame-by-frame analysis of the film. This analysis also revealed transient asymmetries of the face, particularly a transient yet rapid and pronounced strabismus that was not observable at normal projection speeds.

To facilitate analysis of the film, each frame was consecutively numbered from the beginning in order to identify, by means of frame numbers, the location and duration of the derangement of conjugate eye movements. The entire film was scanned by means of a 16-mm. time-and-motion-study projector in which the film transport is operated manually. A frame or series of frames may be repeatedly viewed or contrasted with another frame or series of frames. Thus the frame at which one eye begins to deviate from the axis of the other eye could be noted precisely.

Strabismus

The phenomenon that is herein referred to as "strabismus" occurred as follows. The strabismus occurred over a very short period of time; $\frac{1}{2}$ of a second or less. As an example, the right eye might move to the right for three frames ($\frac{1}{2}$ of a second) while the left eye remained still. The ensuing divergent state is not marked. The movements into and out of divergence were not detectable at normal projection speed. When examined frame by frame, three characteristic types of divergent eye movement emerged.

1. One eye would move horizontally to the left or right while the other remained still. The eye might then shortly move back to its original position or remain in the divergent state. In several instances, for example, the eye would move right for one frame ($\frac{1}{2}$ of a second) and then move back left again on the next frame. This gave the appearance of a rapid "bobble" of the eye.

2. Both eyes diverged. This occurs quite rapidly, often at $\frac{1}{2}$ of a second. For example, the right eye

¹ Requests for reprints should be sent to William S. Condon, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213.

² Obtained through the Psychological Cinema Register, Pennsylvania State University, University Park, Pennsylvania.

will move right and the left eye will move left simultaneously at the same $\frac{1}{4}$ of a second, and only at that $\frac{1}{4}$ of a second.

3. While both eyes are moving in the same direction, one eye will move markedly faster than the other eye. In normal behavior the eyes appear to move together at approximately the same speed. This category of divergence is less precise, however, than the preceding two.

There are two aspects of the strabismus that might be detailed: (a) the actual divergent movements of the eyes and (b) the sustained states that result from these diverging movements.

In a word, the eyes move divergently from each other and may then sustain such states of divergence for varying periods of time. Since the *sustained* divergent states were difficult to detect accurately, the present article will deal primarily with the number of times strabismic *movements* occur with respect to the different personalities. Thus, the actual divergent movement of one eye in relation to the other, the going into or out of divergence, is presented rather than sustained divergent states. The subject's face is often not directed exactly toward the camera, also rendering a reliable assessment of sustained deviance difficult or impossible.

RESULTS

In the beginning of the film it appears as if the patient is shown in her various personalities as background while the therapist narrates the history of the case. Since no explicit identification of the personalities occurred in this section it was not used as a source of data. This section comprised 6,400 frames, and there were only three detectable instances of onset of strabismus.

During the rest of the film, 37,300 frames (some of which comprised titles and lasting $25\frac{1}{2}$ min., 73 instances of transient strabismus were clearly noted. Correlation of these phenomena with the three personalities Eve White, Eve Black and Jane appears in Table 1.

There was also a suggestion of a left and right distribution to the divergent movements of the different personalities. Left-eye movements seem to predominate for Eve White; for example, it is the left eye that moves divergently while the right remains still. Conversely, right-eye movements tend to predominate for Eve Black, while a right-eye bobble seems related to Jane I. Table 2 is an analysis of the types and frequency of strabismic movement for each of the personalities.

TABLE 1
DISTRIBUTION OF STRABISMUS BETWEEN PERSONALITIES

Personality	Frequency of strabismus	Time each personality appears in film
Eve White	11	7 min.
Eve Black	56	9 min.
Jane I	6	4 min.
Jane II	0	$2\frac{1}{2}$ min.
Total	73	

DISCUSSION

There are no references in the literature to transient strabismus as described above. The rapidity and extent of such phenomena raise the issue of whether there had been any reference to this patient's eyes by Thigpen and Cleckley (1957). In their book, *The Three Faces of Eve*, they refer to the patient's suffering from "blinding" headaches and from some of the many visual disturbances often reported in migraine. Referring to Eve Black, they observed that "her eyes dance in such quick spontaneous animation," and later that "her eyes dance with mischief."

Another famous case of multiple personality, Miss Beauchamp, a patient of Morton Prince (1906), was described as having headaches and visual disturbances along with many other symptoms. Sally Beauchamp, the personality almost identical to that of Eve Black, was described as having a peculiar form of anaesthesia in that, with her eyes closed, she felt nothing. Prince likens this

TABLE 2
FREQUENCY OF TYPES OF STRABISMUS

Type	Eve White	Eve Black	Jane I
Left eye moves and right eye is still	7	6	
Left eye moves faster than right eye	2	2	
Right eye moves and left eye is still		29	1
Right eye moves faster than left eye		6	1
Eyes diverge	1	9	
Eyes converge	1	4	
Right-eye bobble			4
Total	11	56	6

anaesthesia to hysterical monocular amblyopia in which a subject cannot see with the blind eye if the other is closed, yet sight returns to the affected eye as soon as the opposite eye is opened. More significant is the statement by Sally that while Miss Beauchamp is concentrating her mind on central vision, she, Sally, need not pay attention to that, but can concentrate her own mind on peripheral vision and recognize things in the periphery.

Anna O., the famous case of Freud's colleague, Breuer (Breuer & Freud, 1953), was described as being split into two personalities, one of which the patient called "her bad self." At the onset of her illness she showed the following visual disturbances: a convergent squint with diplopia; deviation of both eyes to the right so that when her hand reached out for something, it always went to the left of the object.

Such clinical observations suggest the possibility of a more than purely semantic relationship between the dissociation of a personality and the dissociation of normal oculomotor parallelism. A somewhat similar, but less marked, dissociation of normal oculomotor parallelism has also been noted in several films of schizophrenic patients. This is a very tentative finding that will require further study. Strabismus of this form has not been observed thus far in any normal behavior (over 100 films analyzed frame by frame).

The literature relating to oculomotor imbalance indicates that divergence of one eye from its parallel axis with the other eye is varied in its nature. Both systemic and local conditions may affect ocular neuromuscular mechanisms. The oculomotor disturbance may be intermittent, as in the condition known as intermittent exotropia in which the two eyes appear for the most part to be well coordinated, yet one eye may suddenly turn out, often through a rather large angle of divergence. The patient is really not aware that the eye has turned out nor does he experience double vision. The mechanism whereby coordinated binocular vision is interrupted is not known.

REFERENCES

- BREUER, J., & FREUD, S. Studies in hysteria. (Orig. publ. 1895) In, *The standard edition of the complete psychological works of Sigmund Freud*. Vol. 2. London: Hogarth, 1953.
- CONDON, W. S., & OGSTON, W. D. A segmentation of behavior. *Journal of Psychiatric Research*, 1967, 5, 221-235.
- GUTHOR, G. P. *Squint and allied conditions*. New York: Grune & Stratton, 1959.
- PRINCE, M. *The dissociation of a personality*. New York: Longmans, Green, 1906.
- THIGPEN, C. H., & CLECKLEY, H. A case of multiple personality. *Journal of Abnormal and Social Psychology*, 1954, 48, 135-151.
- THIGPEN, C. H., & CLECKLEY, H. *The three faces of Eve*. Kingsport, Tenn.: Kingsport Press, 1957.

(Received August 8, 1968)

TASK DIFFICULTY, RELAXATION, AND ANXIETY LEVEL DURING VERBAL LEARNING AND RECALL¹

JAMES H. STRAUGHAN²

AND

W. HENRY DUFORT

Western Washington State College

Bandon, Oregon

The hypothesis that relaxation instructions help the recall of verbal material by anxious Ss but hinder the recall by nonanxious Ss was tested. High- and low-anxious groups were selected on the basis of their performance on the Anxiety scale of the MMPI. The 121 Ss were presented either 10 easy or 10 difficult paired associates under four conditions of relaxation: (a) no relaxation, (b) relaxation prior to acquisition, (c) relaxation prior to delayed recall, or (d) relaxation prior to both acquisition and delayed recall. The Ss were tested immediately following acquisition and 48 hr. later. Analysis of reciprocal latencies supports the hypothesis. Relaxation was associated with significantly better performance for Ss with high anxiety and poorer performance for Ss with low anxiety.

The relationships between anxiety, relaxation, and certain aspects of learning and performance are relevant topics for investigation because of the continuing interest in desensitization as a therapeutic technique (Paul, 1966; Wolpe, 1958). Current clinical research has been largely directed toward examining the assumption that relaxation is a desirable antidote for a variety of problems involving anxiety, and little attention has been paid to other variables associated with the use of relaxation. Results have generally supported the use of desensitization as opposed to other methods of therapy, for example, test-anxious students report significant reductions of anxiety following desensitization (Emery & Krumboltz, 1967; Garlington & Cotler, 1968).

From the point of view of behavior theory, there is no reason to expect all degrees of relaxation to be equally desirable or to expect overt behavioral changes to be the same for all Ss. Where simple approach-avoidance behavior is involved, any reduction in specific anxiety should permit closer approach to the feared object. Where more complex behaviors are involved, such as with test taking, anxiety may be disruptive if it leads to avoidance or escape responses which interfere with per-

formance or if it reduces the ability to make appropriate discriminations. In general, persons low in anxiety might do better on tasks such as test taking because they are less likely to have developed interfering behaviors. Persons who do perform adequately (mostly those low in anxiety) should do worse following relaxation because their test-taking behavior is associated with a certain level of autonomic feedback. Changing the internal stimulation would reduce the likelihood that adequate test-taking behavior would occur. Persons who are severely anxious during examinations should be better when relaxed because interfering behaviors and excessive stimulation from autonomic feedback are reduced.

Some confirmation is offered for the preceding analysis by research showing that it is useful to distinguish between facilitating and debilitating test anxiety (Alpert & Haber, 1960). An example of an item on the facilitating scale is: "Nervousness while taking a test helps me do better." An item from the debilitating scale is: "Nervousness while taking an exam or test hinders me from doing well." The authors interpret the results to mean that whether anxiety is facilitating or debilitating depends on the kinds of behaviors associated with feedback from anxiety reactions.

Recognizing a distinction between facilitating and debilitating test anxiety would help account for the finding that desensitization

¹ This research was aided by a grant from the Sigma Xi RESA Research fund.

² Requests for reprints should be sent to James H. Straughan, Experimental Education Unit, University of Washington, Seattle, Washington 98105.

TABLE 1

MEAN RECIPROCAL LATENCIES, SECONDS, FOR
CORRECT RESPONSES ON FIRST TEST TRIAL

Group	Hard task		Easy task	
	No relax	Relax	No relax	Relax
High anxiety	.197	.251	.796	.819
Low anxiety	.290	.238	.878	.835

does little to improve test-taking effectiveness, although it does reduce anxiety. Not all anxious students will be helped by a reduction in their anxiety. Neither Garlington and Cotler (1968) nor Emery and Krumboltz (1967) found a significant improvement in grades following desensitization. Garlington and Cotler pointed out that many variables are involved besides test anxiety and give an example of two Ss who lost motivation for study following desensitization. Paul and Eriksen (1964) found that test anxiety was associated with lower scores during examinations but that reducing the anxiety did not raise scores. Paul and Shannon (1966) did find significant differences between experimental and control groups, but the differences seem to have been largely due to a marked drop in the small control group's grade-point average. In each of these studies grade-point average or final examination scores improved for the desensitization groups more than for control groups, suggesting that there is a small effect from relaxation.

A laboratory experiment by Pascal (1949) bears upon questions of the effects of relaxation on test taking. Pascal found that brief relaxation instructions significantly improved the recall of previously learned nonsense syllables. He used several different time intervals between learning and recall and presented the material auditorially to the S, spelling the nonsense syllables. Relaxation immediately prior to recall significantly improved recall but relaxation instructions given immediately after learning was ineffective. His Ss were not selected for anxiety level.

One shortcoming of all of these studies has been the lack of control of relaxation during acquisition. If anxiety interferes with acquisition, the effectiveness of relaxation during recall must be limited. The study to be reported

here throws some additional light on these variables by controlling relaxation during both acquisition and recall in a laboratory learning task. It is predicted that (a) Ss who measure low in anxiety will be hindered by relaxation instructions but that Ss high in anxiety will be helped and (b) relaxation will have these effects during both acquisition and recall. In addition, the present study includes two degrees of task difficulty in line with the suggestion often made that anxiety is more disruptive on more difficult tasks (e.g., Standish & Champion, 1960).

METHOD

Subjects. The Ss were 121 college students who were assigned to one of two groups on the basis of their performance on the Anxiety (A) scale of the Minnesota Multiphasic Personality Inventory (MMPI). The 77 Ss who scored low on the A scale (4 or below) were included in the low-anxiety group, and the 44 Ss who scored high on the A scale (20 or above) were included in the high-anxiety group. The criteria scores for inclusion in these two groups were set on the basis of preliminary tests at a level which seemed likely to insure that approximately equal numbers of Ss would be available for each of the two groups and that these Ss would indeed be from the extremes of the population in terms of anxiety.

Apparatus and materials. A list of 10 easy paired associates was constructed by using common associates in a list prepared by F. R. Fosmire (1962). The association value of these words had been determined by using approximately 430 students at the University of Oregon as Ss during 1960 and 1961. The association value of the words in the easy list ranged from .755 (table-chair) to .474 (butter-bread) with a median of .558. A list of 10 difficult paired associates was constructed by taking 10 other stimulus words from Fosmire's list and selecting for them response words which met two criteria. These criteria were that (a) the word had not been given as a response to the stimulus word by any of Fosmire's Ss and (b) the two Es agreed that the association was a difficult one. As an additional precaution taken in order to insure that the correct word was not initially dominant, stimulus words for the hard list were chosen only if they had responses of high-association value associated with them by Ss in Fosmire's sample. For example, in the hard list the paired associate girl-brick was used. The common response given to girl was boy with an association value of .548. The common responses given to the stimulus words in the hard list had association values ranging from .723 to .426 with a median of .548.

Procedure. The items were presented in a memory drum with each of the 10 items being exposed for 6 sec. with a 6-sec. interval between items. The Ss were given a training trial and two test trials. On the training trial each paired associate was presented

TABLE 2
MEAN RECIPROCAL LATENCIES, SECONDS, OF CORRECT RESPONSES ON SECOND TEST TRIAL

Group	Hard task				Easy task			
	No relax	Relax acquisition	Relax recall	Relax both	No relax	Relax acquisition	Relax recall	Relax both
High anxiety	.262 (8)	.215 (4)	.252 (7)	.218 (6)	.726 (5)	.806 (5)	.789 (4)	.789 (5)
Low anxiety	.309 (10)	.190 (11)	.266 (8)	.141 (10)	.867 (11)	.785 (8)	.751 (9)	.760 (10)

Note.—Numbers in parenthesis are the number of Ss in each group.

to S and on the test trials only the stimulus words were presented. The first test trial followed the training trial by 6 sec. and the second test trial was given 48 hr. later. The *E* measured latencies by means of a stopwatch and recorded results to the nearest tenth of a second.

In addition to the divisions based on anxiety level and task difficulty Ss were divided into four groups on the basis of relaxation instructions. These four groups were: (a) relaxation instructions given immediately before the training trial, (b) relaxation instructions given immediately before the second test trial, (c) relaxation instructions given before both the training trial and the second test trial, and (d) no relaxation instructions. The relaxation instructions were a modified form of those used by Pascal (1949) and required about 5 min. to read to S. The S was asked to concentrate on relaxing major muscle groups one at a time and to coordinate relaxation with exhalation.

RESULTS AND DISCUSSION

Dependent variables for this study are reciprocal latencies (seconds) and number of correct responses. Table 1 presents the mean reciprocal latencies of correct responses on the first test trial. An analysis of variance shows that on the first test trial (a) Ss with low measured anxiety reacted significantly faster than Ss with high measured anxiety ($F = 7.83$, $df = 1/113$, $p < .01$), (b) reaction times were faster for the easy task than for the hard task ($F = 439.96$, $df = 1/113$, $p < .001$), and (c) the effects of the relaxation instructions depended on the measured anxiety level of the Ss; that is, relaxation instructions made the nonanxious Ss slower and the anxious Ss faster ($F = 11.42$, $df = 1/113$, $p < .01$). A significant triple interaction can be interpreted to mean that the effect of the relaxation instructions is somewhat dependent on task difficulty; that is, the effects of the relaxation instructions are greater for the hard task than for the easy task ($F = 3.88$, $df = 1/113$, $p < .01$).

The mean reciprocal latencies of correct response on the second test trial are presented in Table 2. An analysis of variance shows two significant *F*s. First, the latencies for the groups that had the easy task are significantly faster than for the groups that had the hard task ($F = 384.17$, $df = 1/105$, $p < .001$). The mean reciprocal latency for the easy task was .789 and that for the hard task was .230. Second, and more important, there was a significant Measured Anxiety Level \times Relaxation Instructions interaction ($F = 3.75$, $df = 2/105$, $p < .05$). The absence of a significant triple interaction can be interpreted to mean that the double interaction based on the groups that had the harder task is not significantly different from the double interaction based on the groups that had the easy task. The data for the hard and easy tasks were combined into an $A \times B$ table in order to better examine the meaning of the interaction Anxiety Level \times Relaxation. By inspection it appeared that the significant interaction was based on the differences between the groups that did not have relaxation instructions and the groups that did. That this was the case is indicated by the results of *t* tests of the possible subinteractions using the error term from the preceding analysis of variance. The interaction Anxiety Level \times Relaxation corresponds to that of the first test trial; that is, the relaxation instructions slowed the low-anxiety groups and speeded up the high-anxiety groups.³ An additional finding indicated by the *t* tests is that there was a significant interaction between anxiety level and certain of the other conditions of relaxation: (a) between relaxation during re-

³ The high-anxiety groups that had the hard task were made slower by relaxation instructions (Table 2) but not enough so to make the triple interaction significant.

call and relaxation during both acquisition and recall and (b) between relaxation during recall and relaxation prior to acquisition. The meaning of both of these interactions is that relaxation prior to acquisition is more effective than relaxation prior to recall in making the high-anxiety group faster and the low-anxiety group slower.

Another possible analysis of the effects of the independent variables on response latencies during the second test trial involves the changes from Trial 1 to Trial 2. An analysis of the variance in these changes shows no significant *F*s.

When the number of correct responses is considered as the dependent variable, all of the *S*s performed so well on the easy task that no effects of the independent variables can be seen. Only one *S* missed as many as two responses on either the first or second test trial. On the hard task, although the results are in the expected direction with the relaxation instructions making the high-anxiety *S*s better and the low-anxiety *S*s worse, the results are not significant. An indication of the level of difficulty of the hard list can be seen from the fact that the average number correct was 4.484 for the first test trial and 3.766 for the second test trial.

The results of this experiment support the view that relaxation does make a difference in the immediate and delayed recall of verbal material. As predicted, *S*s who measured low in anxiety were hindered by relaxation instructions, but *S*s who measured high in anxiety were helped. These results, when taken together with the results of experiments on test anxiety already cited, offer strong support for the view that desensitization is helpful to many persons who are anxious when taking tests. Under the conditions of this experiment relaxation prior to acquisition was more effective in helping the high anxiety group than relaxation prior to recall. Although an extension of this finding to classroom material is speculative, the obvious interpretation would be that desensitization for test anxiety should also involve the learning situation prior to test taking. The person who becomes less anxious during tests cannot perform better if other anxieties have hindered his learning the material in the first place.

Although it was not anticipated, the present experiment affords an opportunity to re-examine the hypothesis of Kleinsmith and Kaplan's (1960) concerning activation level and consolidation. They found that stimulus words which caused more arousal as measured by the galvanic skin response were conducive to poorer immediate recall of a paired associate but better delayed recall than stimulus words which caused less arousal. In the present experiment the authors have different levels of arousal associated with *S*s rather than with stimulus material. The authors have examined the data on average numbers of correct responses, but find no suggestion that delayed recall is superior for *S*s who are more anxious. As was already stated, reciprocal latencies did not change significantly from Trial 1 to Trial 2.

REFERENCES

- ALPERT, R., & HABER, R. N. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, 1960, 61, 207-215.
- EMERY, J. R., & KRUMBOLTZ, J. D. Standard versus individualized hierarchies in desensitization to reduce test anxiety. *Journal of Consulting Psychology*, 1967, 14, 204-209.
- FOSMIRE, F. R. Differences in associative clustering as a function of associative commonality. Unpublished manuscript, University of Oregon, 1962.
- GARLINGTON, W. K., & COTLER, S. B. Systematic desensitization with test anxious college freshman girls. Paper presented at the meeting of the Washington State Psychological Association, Crystal Mountain, May 1968.
- KLEINSMITH, L. J., & KAPLAN, S. Paired associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, 1960, 65, 190-193.
- PASCAL, G. R. The effect of relaxation upon recall. *American Journal of Psychology*, 1949, 62, 33-47.
- PAUL, G. L. *Insight vs. desensitization in psychotherapy*. Stanford: Stanford University Press, 1966.
- PAUL, G. L., & ERIKSEN, C. W. Effects of test anxiety on "real-life" examinations. *Journal of Personality*, 1964, 32, 480-494.
- PAUL, G. L., & SHANNON, D. R. Treatment of anxiety through systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, 1966, 71, 119-123.
- STANDISH, R. R., & CHAMPION, R. A. Task difficulty and drive in verbal learning. *Journal of Experimental Psychology*, 1960, 59, 361-365.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press, 1958.

(Received September 3, 1968)

RELATIVE EFFECTS OF PROLONGED SOCIAL ISOLATION AND CONFINEMENT: BEHAVIORAL AND EEG CHANGES¹

JOHN P. ZUBEK,² L. BAYER, AND JEAN MARY SHEPHARD

University of Manitoba

This study attempted to determine the relative effects of social isolation and confinement, of 1-wk. duration, on measures of EEG activity, intellectual test performance, affect, and a variety of psychological and somatic symptoms derived from Myers postisolation questionnaire. No significant differences in performance were observed on any of the 11 tests of the intellectual battery or on measures of subjective stress and mood. However, a significant decrease in occipital alpha frequency was obtained in both the socially isolated and confined groups after the 1-wk. period. Since there was no significant difference between these two groups in the magnitude of the decrease, this EEG change appears to have resulted from confinement alone. An analysis of the questionnaire data revealed significant *F* ratios on 16 of the 22 content areas. Ten of the 16 symptom categories were associated with confinement alone and the remainder were associated either with social isolation or a combination of social isolation and confinement. Finally, the results of this study were compared with those reported in several earlier, 1-wk. experiments on perceptual deprivation, a complex condition involving not only social isolation and confinement but also a reduction in sensory stimulation.

In the typical perceptual deprivation (PD) experiment, *S* is exposed not only to a reduction in sensory stimulation but also, of necessity, to social isolation and physical confinement in a small windowless environment. Although various suggestions have been made in the literature that social isolation and confinement may be important factors, particularly in the long-duration experiments, little attention has been directed at dissecting the total PD situation and determining the relative role of its various components. Furthermore, what research is available has only provided an incomplete picture, for example, a comparison of the effects of sensory restriction and social isolation (Persky, Zuckerman, Basu, & Thornton, 1966; Suedfeld, Grissom, & Vernon, 1964; Zuckerman, Albright, Marks, & Miller, 1962; Zuckerman, Persky, Hopkins,

Murtaugh, Basu, & Schilling, 1966) or a comparison of the effects of sensory restriction and a combination of confinement plus restricted motor activity (Zubek & MacNeill, 1967; Zubek & Schutte, 1966). In none of these studies, unfortunately, was an attempt made to determine the specific role of confinement per se.

Recently, however, Zuckerman, Persky, Link, and Basu (1968), in an important extension of their earlier work, analyzed the relative roles of not only sensory restriction, social isolation, and confinement, but also of experimental set and type of *Ss*. Using an 8-hr. duration, measures were taken of two hormone metabolites (17-KGS and 17-KS), time estimation, affect, and responses to a variety of questionnaires containing items specifically designed for conditions of isolation and confinement. Their main conclusion was that "the stress effects of confinement are rather massive and are found even when *Ss* are neither sensorily or socially isolated [p. 194]." In view of these results, it is possible that confinement alone may also be an important factor in producing some of the other effects reported in PD experiments, for example, slowing of electroencephalogram

¹ This study was supported by the Defense Research Board, Canada (No. 9425-08) and by the National Research Council, Canada (APA-290). The authors express their appreciation to F. Greenberg, G. Hughes, S. Milstein, D. Oleson, and M. Selchen for research assistance during the conduct of the experiment.

² Requests for reprints should be sent to John P. Zubek, Department of Psychology, University of Manitoba, Winnipeg, Canada.

(EEG) activity and cognitive deficits (see Zubek, 1969b, for review).

The two-fold purpose of this study is (a) to determine the relative effects of social isolation and confinement, of 1-wk. duration, on measures of EEG activity, cognitive test performance, affect, and various psychological and somatic symptoms elicited from a 242-item questionnaire and (b) to compare these effects with those reported in earlier studies employing 1-wk. of PD (e.g., Zubek, Aftanas, Sansom, Schludermann, Wilgosh, & Winocur, 1962; Zubek & MacNeill, 1967; Zubek & Schutte, 1966; Zubek & Welch, 1963).

METHOD

Subjects

The Ss were 66 male university students who volunteered to spend 1 wk. in the University of Manitoba isolation chamber. None of them had participated in any prior isolation experiment. The remuneration was \$100.00 for the 1-wk. experimentals and \$30.00 for the 1-wk. controls.

Conditions

Confinement group (CF). This group consisted of 21 Ss who were placed individually in an air-conditioned, dome-shaped isolation chamber, measuring 7 ft. in height, 9 ft. in diameter, and 7.5 ft. at the base, for a prescribed period of 1 wk. (for details of chamber see Zubek, Sansom, & Prysiaziuk, 1960). All Ss except 1 (duration = 17 hr.) completed the experiment. Their mean age was 20.2 yr.

In this condition, the inside of the isolation chamber was covered with brightly colored pictures, no restrictions were placed on motor activity (exercising, however, was not permitted), and S had constant access to reading material, a variety of jigsaw puzzles, a radio, a portable television set, a mattress, and a chair with a tablet arm. The lights were put out for 9 hr. at night. In order to avoid social isolation, E visited S at least twice in the morning, afternoon, and evening. Furthermore, two visiting periods per day were scheduled for Ss' friends and acquaintances. Finally, the two doors leading to the chamber were left open enabling S to hear, but not see, any on-going activity in the outside control room area.

Socially isolated group (SI). This group consisted of 25 Ss who were placed individually in the isolation chamber for a 1-wk. period. Of this group, 5 Ss terminated the condition prematurely, quitting after a mean duration of 47 hr. The mean age of the 20 successful Ss was 20.6 yr. The experimental environment was the same as for the CF group with the exception of three modifications designed to achieve a condition of social isolation. First, the radio was replaced with piped in music by Muzak

(exclusively instrumental music). Second, a viewmaster with 300 travel slides was substituted for the television set. Third, no visitors were allowed and conversation between S and E was kept to a minimum (except at the relatively infrequent test sessions). Furthermore, the two doors leading to the chamber were closed at all times.

Ambulatory control group (AC). This group, consisting of 20 Ss (mean age = 20.4 yr.), came to the laboratory at specified intervals during a 1-wk. control period, that is, at the times corresponding to the test sessions of the experimental groups. However, they were free to leave and do what they wanted in the intervening periods.

Measures

Intellectual battery. This battery, developed by Zubek et al. (1960), consists of 11 different tests of a primary mental ability nature, with 8 equivalent forms per test, measuring numerical reasoning, number facility, abstract reasoning, verbal reasoning, verbal fluency, space relations, digit span, recall, recognition, cancellation of numbers, and dexterity. This battery was administered before and at daily intervals during the 1-wk. period. Total administration time was approximately 50 min.

Measures of affect. An appraisal was made of Ss' affective state by two paper-and-pencil tests. The first was the Subjective Stress Scale (SSS) used extensively in experimental situations inducing various degrees of stress (e.g., Berkun, Bialek, Kern, & Yogi, 1962; Myers, 1964). The SSS is a 15-item Thurstone scale, with higher scores indicating more adverse reactions. The second test, developed by Myers, Murphy, Smith, and Goffard (1966) specifically for isolation and confinement experiments, is a measure of overall mood. It consists of a 114-item adjective checklist which S answers as applying "not at all," "somewhat or slightly," or "mostly or generally" to him during a particular time period. Both the SSS and the mood test were administered prior to the experiment and at daily intervals during the 1-wk. period. On the pretest, the Ss were required to answer the items in terms of their experiences occurring during the preceding week of normal life.

Myers postisolation questionnaire. This questionnaire, devised by Myers, Murphy, and Terry (1962) consists of 242 items embracing a variety of typical responses to situations of isolation and confinement. Although scoring keys are provided for 23 categories or content areas, one category (Army assignment) was not employed because of its inappropriateness for university students, as Ss. These 242 items were so devised as to be equally applicable to both experimental and control conditions. The SI and CF groups answered the items as they applied to the week in the isolation chamber; the AC group answered the same items as they applied to the corresponding period of normal life.

EEG activity. The EEGs were obtained from 45 Ss (15 Ss in each group) before and after the 1-wk.

period. The records were taken on an Offner Type T machine at the same time of the day (10:00 A.M.) on both test sessions with precautions to eliminate any drowsiness. Sixteen Ediswan monopad electrodes were applied to the head in International Federation 10/20 positions and two S.L.E. suction electrodes to the ears. Average, bipolar, and monopolar recordings were taken. In order to obtain a quantitative measure of EEG changes, the period-count technique (Engel, Romano, Ferris, Webb, & Stevens, 1944) was used to determine the mean occipital alpha frequency of each S before and after the completion of the 1-wk. period. This involved counting, by means of a Marshall ruler, the number of waves occurring in each of 200, 1-sec. samples of artifact-free occipital lobe tracings and then obtaining the mean frequency per second. In several cases, 200 artifact-free samples could not be obtained and therefore a somewhat smaller sample had to be employed.

RESULTS

EEG activity. The mean posttest change (pre-post difference) in occipital alpha frequency after 1 wk. of the SI, CF, and AC conditions was $-.62$, $-.50$, and $+.04$ cps, respectively. The decrease in EEG frequency in both of the experimental groups was statistically significant (p 's $< .01$), using a t test for correlated measures. Furthermore, 14 of the 15 Ss in each of the two groups exhibited a lower EEG frequency on the postexperimental relative to the preexperimental test period. On the other hand, the AC group, showing virtually a zero change, exhibited a chance distribution of increases and decreases. Finally, an analysis of variance comparing the magnitude of the EEG changes shown by the three groups of Ss, yielded a significant F ratio ($F = 14.29$, $p < .001$). Subsequent t tests revealed that the results of both the SI and CF groups were significantly different from those of the AC group (p 's $< .001$). However, the difference between the SI and CF groups was not significant.

Intellectual battery. An analysis of variance, comparing the relative changes between the three groups of Ss (pre-mean of 7 daily tests), was performed on each of the 11 tests of the intellectual battery. None of the F ratios were significant. An examination of trends, however, revealed that the SI group performed worse than did the AC group on all except 2 of the 11 tests (numerical and verbal reasoning). No such general trend was shown by the CF group: on only 5 of the 11

tests was their performance lower than that of the AC group.

Measures of affect. The results on the SSS and mood scale indicated that both the SI and CF groups experienced a greater degree of subjective stress and possessed a more negative mood than did the AC group during the 1-wk. period. However, on neither of the two affective measures were any of the differences statistically significant.

Myers postisolation questionnaire. The mean scores of the three groups of Ss on the 22 content areas of the retrospective questionnaire are summarized in Table 1. It can be seen that significant F ratios were obtained on 16 content areas. The 6 categories that were not affected were attitude toward Es, religious thoughts, feelings of hunger, tedium, regrets about experimental participation, and feelings of well-being. Application of t tests to the remaining content areas revealed that both the SI and CF groups reported a significantly greater incidence of a variety of phenomena than did the AC group, for example, visual experiences of a hallucinatorylike nature, inefficient thought processes, subjective restlessness, and worry (p 's $< .001$). They also reported a greater frequency of reminiscence and vivid memories, restless acts, and changes in body image and self-appraisal (p 's $< .05$). However, on none of these 8 content categories was there a significant difference between the SI and CF groups. Thus, it would appear that all of these subjective phenomena resulted solely from confinement. Two other phenomena, that is, novel ideas and loss of contact with reality, also appear to be affected to some degree by confinement. Both of these were more prevalent in the CF than in the AC group (p 's $< .01$) and in the SI than in the CF group (p 's $< .05$).

Further analysis of the questionnaire data revealed a significantly greater presence of persistent and vivid dreams, sexual thoughts, and speech difficulties in the SI than in the AC group (p 's $< .01$). None of the other differences were significant, indicating that these effects were related to a combination of social isolation and confinement. Finally, the SI group showed a significantly greater amount of temporal disorientation, feelings of hostility,

TABLE 1
MEAN SCORES OF TWO EXPERIMENTAL AND ONE CONTROL GROUP ON EACH
OF TWENTY-TWO QUESTIONNAIRE CONTENT AREAS

Content areas	Perceptual deprivation ^a N = 18	Social isolation N = 20	Confinement N = 20	Ambulatory controls N = 20	F
Reported visual sensations ("hallucinations")	6.61	5.80	5.00	1.00	14.78***
Dreams	5.39	5.60	4.45	2.45	4.86*
Reminiscence and memory	3.44	3.25	2.85	1.75	6.50**
Sexual thoughts	2.67	2.05	1.55	0.85	5.25**
Novel ideas	2.50	3.15	2.00	0.55	13.09***
Speech difficulties	1.78	1.30	0.95	0.30	4.74*
Self-appraisal	3.11	3.50	2.80	1.50	8.54***
Inefficiencies of thought	14.39	14.40	12.16	6.30	16.40***
Loss of contact with reality	3.83	3.80	2.35	0.65	12.07***
Positive attitude to <i>Es</i>	4.61	3.30	3.45	3.30	0.24
Religious thoughts	2.00	1.45	0.90	0.90	0.93
Loneliness	1.94	2.60	1.30	1.15	8.92***
Hunger	3.22	3.50	2.65	1.95	2.81
Tedium	1.72	1.80	1.25	0.70	2.78
Temporal disorientation	2.78	2.05	1.15	0.85	9.05***
Subjective restlessness	3.50	3.00	3.10	1.00	16.97***
Restless acts	1.94	2.55	2.20	1.50	7.50**
Hostility (anger)	2.56	3.20	2.05	1.60	6.14**
Regret participation	2.17	2.10	1.75	0.85	2.93
Worry	11.22	13.45	12.15	5.25	9.32***
Feelings of well-being	4.94	4.75	4.55	4.20	0.74
Body-image changes	1.72	1.60	1.25	0.30	5.39**

^a For comparative purposes, scores from this previous 7-day perceptual deprivation experiment are also given.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

ity, and loneliness than did either the CF or AC groups (p 's $< .02$). However, on none of these three content areas was there a significant difference between the CF and AC groups. Thus, they appear to be associated solely with social isolation.

Table 1 also shows, for comparative purposes, the mean scores of 18 Ss from an earlier study (Zubek & MacNeill, 1967), who received the Myers questionnaire after completion of a week of PD. It can be seen that the scores of these PD Ss are approximately the same as those of the SI group on most of the 22 content areas. The only areas in which substantially higher scores were obtained were visual experiences of a hallucinatorylike nature, sexual and religious thoughts, speech difficulties, attitude to *Es*, and temporal disorientation.

DISCUSSION

In the light of the Zuckerman et al. (1968) study, indicating that 8 hr. of confinement can produce a variety of stress effects, it is not surprising that a prolonged period of confine-

ment, of 1-wk. duration, should likewise produce significant changes on several physiological and psychological measures. Of these, the slowing of EEG activity is of particular importance since this phenomenon has been reported in numerous PD experiments (see Zubek, 1969a, for review). In the present study both the SI and CF groups showed a significant posttest decrease in occipital alpha frequency ($-.62$ and $-.50$ cps, respectively). Furthermore, since the difference between the two experimental groups was not statistically significant, this EEG effect appears to have resulted from confinement alone. Although a significant change was observed, its magnitude, approximately one-half cycle, is less than that reported by Zubek and Welch (1963) after 1 wk. of PD. Using Engel et al.'s (1944) period-count technique for analysis of the EEG records, the posttest decrease in occipital alpha frequency was -1.21 cps. These results, therefore, appear to suggest that in the total PD situation approximately half of the EEG change is produced by confinement and half by a reduction in sensory

stimulation. Social isolation, apparently, is not a contributing factor.

In view of the relationship claimed to exist between the presence of EEG slowing and deficits on various cognitive and perceptual tasks (Pollack, 1963), significant impairments on the intellectual battery might be expected. This was not the case. Although a general trend toward poorer performance was observed in the SI group, an analysis of variance showed no significant F ratios on any of the 11 tests. Perhaps in the present study the EEG change was not sufficiently great to produce significant impairments on the intellectual battery. Nevertheless, some experimental support for these negative findings has been provided by Smith (1969) who, in a review of the literature on small groups in confinement, concluded that there is very little evidence to indicate that intellectual impairments occur during periods ranging from 5 to 60 days. In the studies cited, the tests involved measures of verbal reasoning, numerical ability, learning, memory, logical reasoning, and perception, that is, tests similar to the authors' intellectual battery.

Although neither social isolation nor confinement produced any intellectual deficits, it is important to note that the PD condition can produce such an effect. Using the same test battery and duration (1 wk.), Zubek et al. (1962) reported a significant impairment on measures of number facility, numerical reasoning, abstract reasoning, space relations, verbal fluency, recognition, cancellation of numbers, and dexterity. These two sets of results, therefore, appear to indicate that in the complex PD condition the restriction of sensory stimulation is the critical factor in producing the various intellectual deficits; social isolation and confinement appear to exert a negligible effect.

The results on the two paper-and-pencil measures of affect indicated that both the SI and CF groups possessed a more negative mood and experienced a greater degree of subjective stress than did the controls. However, none of the differences were statistically significant. These negative results are not surprising since similar results, on these two measures, were obtained in an earlier Manitoba study employing 1 wk. of PD (Zubek &

Schutte, 1966). Although the results on the SSS were not significant, one cannot conclude that the social isolation and confinement conditions were not stressful. The SSS, consisting of only 15 items, may be too crude a measure. The use of other, more sensitive indicators of stress, may have produced a different picture. This hypothesis is supported by the study of Zuckerman et al. (1968) who, using a comprehensive battery of both behavioral and biochemical (17-KGS and 17-KS) indicators of stress reported that the stress effects of confinement were "rather massive." Further evidence in support of the fact that the experimental conditions were stressful, at least for some of the Ss, is indicated by the differential quitting rates. Five percent of the CF group failed to endure the full 1-wk. period in contrast to 20% of the SI group. Moreover, an even higher quitting rate has been reported in 1-wk. PD experiments. In four studies conducted at the Manitoba laboratory (Zubek, 1963, 1964; Zubek et al., 1962; Zubek & Schutte, 1966), the quitting rate has ranged from 25 to 42% ($M = 33\%$). Thus, using an endurance measure, it is clear that the various experimental conditions become increasingly less tolerable as one proceeds from confinement, to social isolation, and finally to perceptual deprivation.

An analysis of the results on the Myers postisolation questionnaire revealed a variety of unusual psychological symptoms, phenomena similar to those first reported in the McGill PD experiments (e.g., Bexton, Heron, & Scott, 1954) and subsequently confirmed at a number of different laboratories (see Zuckerman, 1964, for review). Of the 16 content areas that showed significant F ratios, 8 were associated solely with confinement, viz., experiences of a hallucinatorylike nature, inefficient thought processes, subjective restlessness, restless acts, worry, reminiscence and vivid memories, and changes in body image and self-appraisal. A further two symptoms (novel ideas and loss of contact with reality) also resulted to some degree from confinement but their incidence was significantly increased by the addition of social isolation. Thus, it is clear that most of the 16 symptoms were affected by confinement. On the remaining content areas, 3 (temporal

disorientation, feelings of hostility, and loneliness) were associated solely with social isolation and another 3 (persistent and vivid dreams, sexual thoughts, and speech difficulties) were related to a combination of social isolation and confinement. Finally, a comparison of these results with those of an earlier 1-wk. PD experiment (Zubek & MacNeill, 1967), using the same questionnaire, revealed that the scores of the PD group were approximately the same as those of the SI group on most of the 22 content areas. The only symptoms in which the PD group showed substantially higher scores were experiences of a hallucinatorylike nature, temporal disorientation, speech difficulties, sexual and religious thoughts, and attitude to *Es*. The incidence of these phenomena, therefore, appears to be increased to some degree by the imposition of a reduction in sensory stimulation.

Although Zuckerman et al. (1968) also used the Myers questionnaire in the only other relevant study, their results cannot be directly compared with the present authors' since the duration involved was only 8 hr. Furthermore, two Ss were placed in their confinement condition, a situation which, according to Smith (1969), sometimes leads to "difficult interpersonal problems." Nevertheless, the results show considerable agreement. Even with their short duration, Zuckerman et al. observed that the effects of confinement were quite general, showing an increased incidence of most of the questionnaire-elicited symptoms. In addition, they also reported that the effects of sensory restriction "tend to be more specific, consisting of an anxiety reaction which is related to the cutting off of one's normal sensory ties with reality and the appearance of unusual perceptions and ideas [p. 194]."

In interpreting the results of this experiment, two factors should be taken into consideration. The first is the physical dimension of the chamber. The Manitoba isolation chamber is quite small, consisting of only 45 sq. ft. of floor space, and in view of this it is possible that the reported results may be related, to some degree, to the use of a very restricted physical environment. In this regard, it is interesting to note that two Ss

who complained about being "hemmed-in" showed the largest posttest decrease in alpha frequency (-1.00 and -1.18 cps) of the CF group, indicating that the size of the living space may be an important variable. Further research, systematically varying the size of the confinement area (both physical and perceived space) might prove fruitful. Second, in the usual PD condition Ss are provided with little or no cues to the passage of time. This was not the case in the present experiment. In order to achieve an essentially normal, nonboring sensory environment during a prolonged period of time, the lights were put out at night, and in the CF condition, a radio and television set were provided, all of which provided Ss with time cues. The presence of this information, therefore, may have diminished somewhat the extent of the social isolation and confinement effects. Despite this, their effects were still considerable.

REFERENCES

- BERKUN, M. M., BIALEK, H. M., KERN, R. P., & YOGI, K. Experimental studies of psychological stress in men. *Psychological Monographs*, 1962, 76(15, Whole No. 534).
- BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology*, 1954, 8, 70-76.
- ENGEL, G. L., ROMANO, J., FERRIS, E. B., WEBB, J. P., & STEVENS, C. D. A simple method of determining frequency spectrums in the electroencephalogram. *Archives of Neurology and Psychiatry*, 1944, 51, 134-146.
- MYERS, T. I. Studies of stressfulness of sensory deprivation and confinement. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1964.
- MYERS, T. I., MURPHY, D. B., SMITH, S., & GOFFARD, S. J. *Experimental studies of sensory deprivation and social isolation*. (HumRRO Tech. Rep. 66-8) Alexandria, Va.: Human Resources Research Office, 1966.
- MYERS, T. I., MURPHY, D. B., & TERRY, D. F. The role of expectancy in subjects' responses to sustained sensory deprivation. Paper presented at the meeting of the American Psychological Association, St. Louis, September 1962.
- PERSKY, H., ZUCKERMAN, M., BASU, G. K., & THORNTON, D. Psychoendocrine effects of perceptual and social isolation. *Archives of General Psychiatry*, 1966, 15, 499-505.
- POLLACK, M. Psychological changes with induced EEG slowing. *American Psychologist*, 1963, 18, 443. (Abstract)

- SMITH, S. Studies of small groups in confinement. In J. P. Zubek (Ed.), *Sensory deprivation: Fifteen years of research*. New York: Appleton-Century-Crofts, 1969.
- SUEDFELD, P., GRISSOM, R. J., & VERNON, J. The effect of sensory deprivation and social isolation on the performance of an unstructured task. *American Journal of Psychology*, 1964, 77, 111-115.
- ZUBEK, J. P. Counteracting effects of physical exercises performed during prolonged perceptual deprivation. *Science*, 1963, 142, 504-506.
- ZUBEK, J. P. Behavioral changes after prolonged perceptual deprivation (no intrusions). *Perceptual and Motor Skills*, 1964, 18, 413-420.
- ZUBEK, J. P. Physiological and biochemical effects. In J. P. Zubek (Ed.), *Sensory deprivation: Fifteen years of research*. New York: Appleton-Century-Crofts, 1969. (a)
- ZUBEK, J. P. (Ed.) *Sensory deprivation: Fifteen years of research*. New York: Appleton-Century-Crofts, 1969. (b)
- ZUBEK, J. P., ARTANAS, M., HASEK, J., SANSOM, W., SCHLUDERMANN, E., WILGOSH, L., & WINOCUR, G. Intellectual and perceptual changes during prolonged perceptual deprivation: Low illumination and noise level. *Perceptual and Motor Skills*, 1962, 15, 171-198.
- ZUBEK, J. P., & MACNEILL, M. Perceptual deprivation phenomena: Role of the recumbent position. *Journal of Abnormal Psychology*, 1967, 72, 147-150.
- ZUBEK, J. P., SANSOM, W., & PRYSIAZNIUK, A. Intellectual changes during prolonged isolation (darkness and silence). *Canadian Journal of Psychology*, 1960, 14, 233-243.
- ZUBEK, J. P., & SCHUTTE, W. Urinary excretion of adrenaline and noradrenaline during prolonged perceptual deprivation. *Journal of Abnormal Psychology*, 1966, 71, 328-334.
- ZUBEK, J. P., & WELCH, G. Electroencephalographic changes after prolonged sensory and perceptual deprivation. *Science*, 1963, 139, 1209-1210.
- ZUCKERMAN, M. Perceptual isolation as a stress situation: A review. *Archives of General Psychiatry*, 1964, 11, 255-276.
- ZUCKERMAN, M., ALBRIGHT, R. J., MARES, C. S., & MILLER, G. L. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychological Monographs*, 1962, 76(30, Whole No. 549).
- ZUCKERMAN, M., PERSKY, H., HOPKINS, T. R., MURTAUGH, T., BASU, G. K., & SCHILLING, M. Comparison of stress effects of perceptual and social isolation. *Archives of General Psychiatry*, 1966, 14, 356-365.
- ZUCKERMAN, M., PERSKY, H., LINK, K. E., & BASU, G. K. Experimental and subject factors determining responses to sensory deprivation, social isolation, and confinement. *Journal of Abnormal Psychology*, 1968, 73, 183-194.

(Received September 10, 1968)

ERRATUM

In the article, "Self-Verbalization and Emotional Arousal," by David C. Rimm and Stuart B. Litvak, which appeared in the April 1969 issue of the *Journal of Abnormal Psychology*, the last sentence in the first paragraph, second column, page 186, should read "For the highly verbal college student, however, by a process of higher order conditioning and additional direct conditioning, objective statements could well become just as affect inducing." The second paragraph, fourth sentence should read "It is reasonable to suppose that once a verbal mediator comes to elicit negative affect or drive, this will motivate instrumental responding." Sixth sentence, "One might predict that when such a sequence elicited negative affect, this would function as punishment suppressing the particular sequence." Seventh sentence, "Suppose, however, that the negative affect did drive an instrumental response, culminating in drive reduction."

MEASURING IMPULSE CONTROL IN INSTITUTIONALIZED DELINQUENTS USING RORSCHACH CONTENT AND THOUGHT PROCESS SCALES¹

ALAN H. ROBERTS²

University of Minnesota

AND

ROBERT V. ERIKSON

Loysville Youth Development Center, Pennsylvania

Institutionalized delinquent males were separated by judges into a more controlled (GGs) and more impulsive (SSs) group. Ss were matched by age, IQ, and ethnic origin. Rorschachs were individually administered and scored on six content scales and two thought process scales. A single control score derived from these scales significantly differentiated the two groups. In particular, the SSs produced more content similar to that hypothesized for manic patients while the GGs demonstrated more thinking which involved exaggerated attempts at control.

Previous studies by Erikson and Roberts (1966) and Roberts and Erikson (1968) have demonstrated that two groups of institutionalized delinquent boys can be reliably differentiated with respect to foresight and planning ability, impulse control, willingness or ability to delay gratification, and ratings of adjustment in a training school situation. The goal of predicting these variables in delinquents is an important one; when prediction is possible, there is a rational basis for planning and counseling.

The purpose of the present study is to determine whether impulsive and controlled delinquents can be differentiated by the Rorschach content and thought process scales of Bower, Testin, and Roberts (1960). An analysis of their intercorrelational and factor-analytic data suggested that eight of their scales could be combined into two sets of four scales each, one set associated with "impulsiveness" (their Catatonic, Manic, and Character Disorder Content scales and the Disorganized Thought Process scale) and another set associated with "control" (their Paranoid, Depressive, and Obsessive-Compul-

sive Content scales and the Arbitrary Tightness Thought Process scale). It is hypothesized that the more impulsive of the two delinquent groups (hereafter termed SSs) will obtain higher scores on the Catatonic, Manic, and Character Disorder Content scales and the Disorganized Thought Process scale while the more controlled group (hereafter termed GGs) will obtain higher scores on the Paranoid, Depressive, and Obsessive-Compulsive Content scales and the Arbitrary Tightness Thought Process scale.² On the basis of this hypothesis, a single control score was derived as described below.

METHOD

A group of 18 SSs and 29 GGs were selected by the staff of the school and from these two groups, 10 pairs were matched by age, IQ as measured by the California Short Form Test of Mental Maturity, and ethnic origin. There were 8 Spanish-American pairs and 2 Anglo-American pairs. The mean age for the SSs was 16.5 ($SD = 1.37$) and the mean IQ was 88.2 ($SD = 12.51$). For the GGs, the means were 16.6 ($SD = 1.19$) and 88.5 ($SD = 9.65$), respectively. The two groups did not differ significantly with respect to length of stay at the training school, number of previous commitments, or grade in school.

¹ This study was supported in part by the Vocational Rehabilitation Administration (Training Grant RT-2). The authors thank the staff of the New Mexico Boys' School for their cooperation in this study and Peter Briggs, Paul E. Meehl, William Schofield, and Auke Tellegen for critically reading an earlier version of this paper.

² Requests for reprints should be sent to Alan H. Roberts, Department of Physical Medicine and Rehabilitation, Mayo Memorial Building, University of Minnesota, Minneapolis, Minnesota 55455.

³ The designations, GGs and SSs, are derived from the vernacular of the training school where staff and students refer to the well-behaved students as "good guys" and the poorly behaved students as "shit sticks." The boys referred to as GGs are more often described by their dormitory supervisors with terms like causes no trouble, quiet, obeys the rules and regulations, etc., while the boys referred to as SSs are more often described with terms like fighting, mouthing off, general nuisance, etc.

TABLE 1

INTERSCORER RELIABILITY AND COMPARISON OF SSs AND GGs ON CONTROL SCORE, CONTENT SCALES, AND THOUGHT PROCESS SCALES

Measure	SSs		GGs		<i>t</i>	Reliability
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Control score	11.7	5.80	18.1	6.39	2.31*	—
Control variables						
Paranoid	2.4	2.95	3.0	3.30	0.97	.69
Depressive	3.2	2.82	5.1	4.33	1.01	.88
Obsessive-Compulsive	0.8	0.92	0.6	0.97	-0.51	.64
Arbitrary Tightening	2.3	2.31	5.7	4.00	3.64**	.71
Impulse variables						
Catatonic	1.8	2.78	0.6	1.07	1.31	.84
Manic	2.6	2.22	1.2	1.93	2.41*	.81
Character Disorder	2.4	2.37	1.6	1.17	1.06	.79
Disorganization	1.5	2.72	1.8	3.05	-0.20	.58

Note.—All scores except control score are in raw score form. *N* = 10 in each group; SS = impulsive group; GG = controlled group.

**p* < .05, one-tailed.

***p* < .01, one-tailed.

Rorschachs were administered individually and the two groups did not significantly differ with respect to the total number of Rorschach responses given. For the SSs the mean was 12.1 (*SD* = 4.65) and for the GGs the mean was 11.5 (*SD* = 5.25). The instructions of Bower, Testin, and Roberts (1960) were used in scoring content and thought process scales. These scales were scored independently by two judges and their interscorer reliabilities are shown in Table 1.

All scale scores were transformed to standard scores combining the two groups for this purpose. The standard scores of scales hypothesized to be associated with control were summed without weighting them, as were the scores hypothesized to be associated with impulsiveness. The impulsiveness total was then subtracted from the control total and a constant was added to each score to eliminate negative numbers. This final figure represented a control score with the higher values associated with greater control.

RESULTS AND DISCUSSION

The mean control scores for SSs and GGs are shown in Table 1. The mean control score for GGs is significantly higher than for SSs supporting the overall hypothesis.

The raw score means for the various individual scales are also compared in Table 1 in order to get a better idea of which scales contribute most to this finding. All means are in the predicted direction except the Obsessive-Compulsive and Disorganization

scales. Of the control variables, the Arbitrary Tightening scale significantly differentiates the two groups while among the impulse variables, the Manic Content scale significantly differentiates the two groups.

The predictive efficiency of the control score, using an ad hoc "best cut," was examined. Maximum separation is achieved with a cutting score of 16. Using this score, 80% of SSs achieve a score of 16 or less while 70% of GGs achieve a score of 17 or more. Twenty percent of the SSs and 30% of the GGs are misclassified with this cut. While some information concerning the base rates of SSs and GGs is available (Roberts & Erikson, 1968), the application of Bayes' rule for calculating inverse probability (Meehl, 1956) may be questionable because of the use of matched pairs. While such an application would be "favorable" in this instance (i.e., assertions based on the score would be more likely to be correct than incorrect), these findings must be cross-validated because of the ad hoc determination of the cutting score.

The results of this study demonstrate that impulsive and controlled delinquent Ss can be differentiated by content and thought process scoring of the Rorschach. In particular, the more impulsive group appears to produce

content similar to that hypothesized for manic patients while the more controlled group tends to demonstrate thinking which involves exaggerated attempts at control.

REFERENCES

BOWER, P. A., TESTIN, R., & ROBERTS, A. Rorschach diagnosis by a systematic combining of content, thought process, and determinant scales. *Genetic Psychology Monographs*, 1960, 62, 105-183.

ERIKSON, R. V., & ROBERTS, A. H. A comparison of two groups of institutionalized delinquents on Porteus Maze Test performance. *Journal of Consulting Psychology*, 1966, 30, 567.

MEEHL, P. E. Wanted—a good cookbook. *American Psychologist*, 1956, 11, 263-272.

ROBERTS, A. H., & ERIKSON, R. V. Delay of gratification, Porteus Maze Test performance, and behavioral adjustment in a delinquent group. *Journal of Abnormal Psychology*, 1968, 73, 449-453.

(Received September 11, 1968)

Manuscripts Accepted for Publication in the *Journal of Abnormal Psychology*

Galvanic Skin Reactivity among Family Groups Containing Disturbed Adolescents: Michael J. Goldstein,* Eliot H. Rodnick, Lewis L. Judd, and Edward Gould: Department of Psychology, University of California, Los Angeles, California 90024.

Paradoxical Enhancement of Learned Fear: Michael Rohrbaugh* and David C. Riccio: Department of Psychology, Kent State University, Kent, Ohio 44240.

Role of Verbal Intelligence in the Field Approach of Selected Groups of Psychotics: Barbara J. Powell*: Department of Psychology, Malcolm Bliss Mental Health Center, 1420 Grattan Street, St. Louis, Missouri 63104.

Changes in Locus of Control as a Function of Life Crisis Resolution: Ronald E. Smith*: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Effects of Sensory Restriction on Susceptibility to Hypnosis—a Hypothesis and More Preliminary Data: Ian Wickramasekera*: Mental Health Clinic, 320 East Armstrong, Peoria, Illinois 61603.

Communication of Intents among Parents and Their Disturbed Adolescent Child: Sigrid McPherson*: Department of Psychology, University of California, Los Angeles, California 90024.

Social Competence and Psychiatric Symptomatology in Japan: A Cross-Cultural Extension of Earlier American Findings: Juris G. Draguns,* Leslie Phillips, Inge K. Broverman, and William Caudill: Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

Absolute Auditory Threshold and Threshold of Unpleasantness of Chronic Schizophrenic Patients and Normal Controls: Fredric M. Levine* and Norman Whitney: Department of Psychology, State University of New York, Stony Brook, Long Island, New York 11790.

Modification of Performance on the Rod-and-Frame Test through Token Reinforcement Procedures: Loring W. McAllister*: Western Mental Health Center, Inc., 438 West Main Street, Marshall, Minnesota 56258.

Preference Intransitivity as an Indicator of Cognitive Slippage in Schizophrenia: Gordon A. Braatz*: Psychology Service, Veterans' Administration Hospital, 54th Street and 48th Avenue South, Minneapolis, Minnesota 55417.

Case Report: The Use of Nonspecific Underlying Motivational Factors in the Systematic Desensitization of Specific Marital and Interpersonal Fears: J. Sidney Shrauger* and Edward S. Katkin: Psychological Clinic, Department of Psychology, State University of New York, Buffalo, New York 14214.

Critique of "Treatment of Insomnia by Relaxation Training": Relaxation Training Rogerian Therapy, or Demand Characteristics?: Russell Eisenman*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

Effect of Subliminal Shock Conditioning on Recall: Robert H. Harrison*: Boston City Hospital, 818 Harrison Avenue, Boston, Massachusetts 02118.

Dichotic Stimulation as a Method of Assessing the Disorder of Attention in Overinclusive Schizophrenic Patients: Robert W. Payne,* Arthur C. Hochberg, and David V. Hawks: Department of Behavioral Science, Temple University, % Eastern Pennsylvania Psychiatric Institute, Henry Avenue, Philadelphia, Pennsylvania 19129.

Contingency Control and Shaping Procedures to Institute Bed-Making Behaviors in a State Mental Hospital: Richard Wolf* and Dan Perkins: Peoria State Hospital, Peoria, Illinois 61607.

Nonfluency, Information, and Word Length: Richard I. Lanyon* and Daryl A. Duprez: Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

Reality Closeness, Phantasy, and Autokinesis: A Dimension of Cognitive Style: Martin Mayman* and Harold M. Voth: Psychological Clinic, 1027 East Huron Street, Ann Arbor, Michigan.

* Asterisk indicates author for whom address is supplied.

PERGAMON

THE DISCRIMINATION PROCESS AND DEVELOPMENT

By J. FELLOWS, *Portsmouth College of Technology*

Provides a comprehensive picture of the various operations involved in the process of visual discrimination and traces the development of these operations in young children. Partial contents include: Classical theories of discrimination learning; Attention, strategies and age; Orienting and planning; A model of matching and discrimination performance; and, Learning and perception of orientation.

1968

230 pp.

\$10.00

THE PRACTICE OF BEHAVIOR THERAPY

By Professor JOSEPH WOLPE, M.D., *Temple Univ. School of Medicine*

This significant book gives a comprehensive coverage of behavior therapy of neuroses, together with the most recent developments in therapy techniques. Professor Wolpe focuses on a practical approach to the neurotic patient and views him in his totality: an organism influenced by his environment, his constitution and his learning. It contains a detailed discussion of techniques to help patients overcome neurotic anxiety and other neurotic habit patterns. Also contains chapters on the use of drugs in behavior therapy and on operant conditioning methods.

1969

326 pp.

\$8.95

MENTAL HEALTH COUNSELORS AT WORK

By THOMAS M. MAGOON *et al.*, *Univ. of Maryland*

Covers the development of roles, utilization of mental health counselors by several agencies, problems encountered, and reports on many evaluations of their performance in schools, clinics, hospitals, and other settings.

1969

914 pp.

\$8.50

THE GENESIS OF THE CLASSICAL CONDITIONED RESPONSE

By IRENE MARTIN, *Institute of Psychiatry, London* and A. B. LEVEY, *Graylingwell Hospital, Chichester*

Describes recent developments in the field of eyelid and autonomic conditioning studies with the new and original method of analysis of conditioning data. Presents a theoretical model which elucidates aspects of conditioning not explained by older theories.

1969

164 pp.

\$9.00

RORSCHACH WITH CHILDREN

By JESSIE FRANCIS-WILLIAMS

Demonstrates the use of the Rorschach Test as a tool of diagnosis in clinical work with children and examines other projective techniques in comparison. Gives a brief description of the underlying theory of the concept of projection together with some account of the kinds of projective techniques used with children.

1968

168 pp.

\$7.00

FOUNDATIONS OF CHILD PSYCHIATRY

Edited by EMANUEL MILLER

Illustrates the present attitudes to theory, method and the basic clinical manifestations and such social issues which seem to play so important a part in the lives of parents and children. Social studies have been given special weight in order to stress the part that the family and the social situation play in the emergence of psychological disorders and the need for a social orientation on therapy.

1968

716 pp.

\$50.00



PERGAMON PUBLISHING COMPANY

A Division of Pergamon Press, Inc.

MAXWELL HOUSE • FAIRVIEW PARK • ELMSFORD, NEW YORK 10523

New

Volume I

***The Measurement of Psychological States
Through the Content Analysis of Verbal Behavior***

Volume II

***Manual of Instructions for Using the Gottschalk-Gleser Content Analysis
Scales: Anxiety, Hostility, Social Alienation-Personal Disorganization***

Louis A. Gottschalk and Goldine C. Gleser

A major need in psychiatric and psychological research has been the development of methods for the accurate measurement of transient emotional, motivational, and cognitive states. The present volumes offer such a method based on language as the raw material from which measurements can be made. Drawing on psychoanalysis, learning theory, and linguistics, the authors present a series of scales, which include categories of crucial verbal expressions, by means of which non-professional technicians can objectively score the content of verbal behavior as it is expressed in a five-minute verbal sample. The magnitude of various mental states, such as anxiety, hostility, cognitive and intellectual impairment, can thus be determined.

1969 LC: 69-15828 Vol. I 317 pages \$12.50
Vol. II 176 pages \$6.00



from California

University of California Press • Berkeley 94720

Oxford University Press



A Textbook of Psychosexual Disorders

SECOND EDITION

By CLIFFORD ALLEN, M.D., *Consultant Psychiatrist*. The thoroughly revised edition of this text for both the general practitioner and specialist discusses eight major topics: The Sexual Instinct and its Development; Disorders of Sexual Expression; Disorders of the Instinctual Object; Disorders of the Sexual Stimulus; Disorders of the Instinctual Strength; Prostitution; The Prevention, Treatment, and Prognosis of the Psychosexual Disorders; The Medico-Legal Aspects of Psychosexual Disorders. References revised and updated.

1969

460 pp.

\$13.75

Henderson and Gillespie's Textbook of Psychiatry

TENTH EDITION

Revised by I. R. C. BATCHELOR, *University of Dundee*. In the completely revised edition of this text for students and practitioners, the authors present psychiatry as an essential component of general medicine, stressing the immense medico-social problems with which psychiatrists must deal. Included is a clarification of affective, paranoid and organic psychoses as well as new data concerning aetiology of mental illness, its treatment, and problems of rehabilitation.

1969

640 pp.

cloth \$13.75

paper \$8.50

Disordered Behavior:

BASIC CONCEPTS IN CLINICAL PSYCHIATRY

By ERIC PFEIFFER, M.D., *Duke University Medical Center*

1968

224 pp.

cloth \$5.75

paper \$4.00

The Experience of Anxiety:

A CASEBOOK

By MICHAEL J. GOLDSTEIN and JAMES O. PALMER, *both of the University of California, Los Angeles*

1963

304 pp.

paper \$2.50

Principles of Group Treatment

By ERIC BERNE, M.D., *McAuley Neuropsychiatric Institute, San Francisco*

1966

384 pp.

\$7.00

Psychological Diagnosis in Clinical Practice:

WITH APPLICATIONS IN MEDICINE, LAW, EDUCATION, NURSING, AND SOCIAL WORK

By BENJAMIN POPE, *The Sheppard and Enoch Pratt Hospital*, and WINFELD H. SCOTT, *National Institute of Mental Health*

1967

360 pp.

18 illus. (10 in color)

\$8.50

OXFORD UNIVERSITY PRESS

200 Madison Avenue, New York, N.Y. 10016

NEW IN 1969

DEVELOPMENTAL PSYCHOLOGY

Editor: Boyd R. McCandless

Developmental Psychology contains articles which represent the broad range of growth and development and their major associated variables. Chronological age as well as sex, socioeconomic status and efforts of physical growth variables (including time of maturation and body build) are all considered relevant development variables.

Although the adolescent and the aging population appear to deserve particular attention at this time, *Developmental Psychology* indeed covers the entire span. Cross-species articles and articles concerning developmental research with retardates are also included.

The articles in *Developmental Psychology* are not limited to any one approach such as "experiments versus studies" or "cognition versus motivation".

**Bimonthly. One volume per year.
Subscription Price: \$10.00 per year**

**AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036**

270
12. 6. 71
Lib

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute
and Columbia University*

Associate Editors: WILLIAM N. THETFORD, *College of Physicians and Surgeons,
Columbia University*

ARTHUR C. CARR, *New York State Psychiatric Institute and Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

RICHARD L. BLUMENTHAL
*New York State Psychiatric Institute
and Columbia University*

EUGENE I. BURDOCK
New York University

JOHN M. BUTLER
University of Chicago

LOREN J. CHAPMAN
University of Wisconsin

W. CRAWFORD CLARK
*New York State Psychiatric Institute
and Columbia University*

JOHN H. CUMMING
*New York State Department of Mental
Hygiene*

W. GRANT DAHLSTROM
University of North Carolina

JOSEPH L. FLEISS
*New York State Department of Mental
Hygiene and Columbia University*

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MARGARET GARNER
University of Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Michigan

ERNEST R. HILGARD
Stanford University

LAWRENCE C. KOLB
*New York State Psychiatric Institute
and Columbia University*

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
United States Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

MARTIN T. ORNE
*Institute of the Pennsylvania Hospital
and the University of Pennsylvania*

ERNST PRELINGER
Yale University

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

JACQUES RUTSCHMANN
*New York State Psychiatric Institute
and Columbia University*

KURT SALZINGER
*New York State Psychiatric Institute
and Brooklyn Polytechnic Institute*

JOSEPH SCHACHTER
*New York State Psychiatric Institute
and Columbia University*

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

JOSEPH ZUBIN
*New York State Department of Mental
Hygiene and Columbia University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (a) psychopathology—its development, treatment or remission, and its symptomatology, and course; (b) normal processes in abnormal individuals; (c) pathological or atypical features of the behavior of normal persons; (d) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (e) social or group effects on adjustment and pathological processes; (f) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is \$10.00 annually (foreign, \$10.50). Single issues are \$2.00.

Because of the large number of manuscripts submitted, a publication lag of up to 24 months can occur. Authors may avoid this delay if they are prepared to pay the costs of early publication (\$45.00 per Journal page).

Manuscripts must be submitted in duplicate and should be sent to the Editor-elect:

Donald R. Peterson
Journal of Abnormal Psychology
Children's Research Center
University of Illinois
Champaign, Illinois 61820

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR	MYRTLE HENDRICKS	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1969 by the American Psychological Association, Inc.

REALITY CLOSENESS, PHANTASY, AND AUTOKINESIS: A DIMENSION OF COGNITIVE STYLE¹

MARTIN MAYMAN²

University of Michigan

AND

HAROLD M. VOTH

The Menninger Foundation, Topeka, Kansas

This study examines the validity and predictive power of an explanatory hypothesis for the occurrence of the autokinetic phenomenon. It assumes that in the autokinetic situation, as in a wide range of life situations, some people characteristically maintain a "reality-close" life style, while others characteristically allow attention to drift toward more subjective endopsychic experiences. The relationship between reality closeness, fantasy proneness, and autokinesis is examined in a pilot study and two replication studies. The relevance of these findings to the psychoanalytic theory of ego functioning is discussed.

Almost anyone who looks steadily at a fixed pinpoint source of light in otherwise absolute darkness is likely soon to see the stationary light begin to "move." The earliest explanations assumed that unconscious eye movements caused retinal shifts which would account for the kinds of light movements reported by Ss, but this belief was easily discounted by a variety of experiments (Royce, Carran, Aftanas, Lehman, & Blumenthal, 1966). Ocular-retinal explanations were soon replaced by speculations that the illusion can be explained in cortical neurophysiological terms, but all such explanations have lacked substance and failed to generate testable hypotheses (Reinwald, 1952).

Contrasting with such simple neurophysiological explanations of the illusion is a growing body of evidence that the illusion is directly responsive to a wide range of experimental conditions affecting the psychological or mental state of the person in the autokinetic test situation (Santos, Farrow, & Haines, 1965; Schwartz & Shagass, 1960; Voth, 1941; Voth & Mayman, 1963). Coincident with these studies has been the emergence of equally impressive evidence that the autokinetic illusion is directly related to central personality

variables. Test-retest studies have shown high repeat reliability in the amount of movement seen (summarized in Voth & Mayman, 1963). Other investigators have shown that autokinesis covaries with personality style (Bush, Hatcher, & Mayman, 1969; Temerlin, 1956; Voth & Mayman, 1963), with clinical behavior (Voth & Mayman, 1966, 1967, 1968), and with choice of illness (A. C. Voth, 1947; H. M. Voth, 1962).

The present study was undertaken in an attempt to further clarify the relationship between autokinesis and personality style by turning attention to psychological processes which might account for individual differences in performance on the autokinetic test.

In previous studies an intervening variable was formulated to help explain the link between such seemingly disparate phenomena as autokinesis and character style.³ It was suggested that persons habitually differ with respect to their "closeness to" or "distance from" their concrete stimulus field. The more "reality close" the person, the less autokinesis will he experience. On the other hand, the more distance a person characteristically maintains from people and things about him, that is, the more "reality distant" the person, the more autokinetic movement does he see.

In this study, the authors attempted to account for this relationship between autokinesis and life style by showing that a significant variable in autokinetic movement perception is S's unintentional retention of an image of the darkened room during the auto-

¹This research was supported by Public Health Service Research Grant No. MH 07547-03, from the National Institute of Mental Health. The authors are much indebted to Phyllis Levy and Marilyn Klein Silverman for their indispensable help in carrying out the study.

²Requests for reprints should be sent to Martin Mayman, Psychological Clinic, 1027 East Huron Street, Ann Arbor, Michigan.

³See especially Voth and Mayman, 1963 and 1966.

kinetic test. The authors tried to show that reality-close people "see little or no autokinetic movement because they retain a firm and lasting impression of the autokinetic room in the dark and retain, thereby, an orienting frame of reference that prevents the occurrence of the illusion. Conversely, the more reality distant the person, the quicker the experimental room fades out of awareness as a subjectively real presence and, consequently, the quicker, more compelling, and more extensive will be his experience of autokinesis.

The authors also attempted to confirm the corollary hypothesis that those persons who are characteristically outwardly rather than inwardly oriented, and whose attention remains outwardly directed even during a fairly extended period of social and perceptual isolation, are less likely to experience imagery and phantasy in the autokinetic situation. On the other hand, the more reality distant a person, the more vivid, rich, and compelling will be the coincident imagery and phantasy which intrude and merge with his autokinetic experience.

In short, this study tested two predictions: (a) Autokinesis will be positively correlated with the extent to which the experimental room fades out of awareness, and (b) The intrusion of imagery and phantasy into the autokinetic experience will be directly related to reality closeness, reality distance as measured (a) by the autokinetic score, and (b) by the degree to which the room fades out of awareness in the course of the experiment.

METHOD

The *S* was seated at a table 9 ft. away from a small aluminum box mounted on a tripod stand which was clearly stationary. The aperture for the light source was .0280 in. in diameter. A $\frac{1}{4}$ -in. photographic filter inside the box produced diffused light, .0006 ftc. in intensity. Before *S* on a table was a drawing board on which was fastened a sheet of newsprint paper 22 × 28 in. The *S* was told he would be left alone in a totally darkened room except for the pinpoint of light which he was to look at steadily for 10 min. He was told that the light may or may not appear to move. Should it move, he was to trace its path with a pencil. If the light at any time

* In previous papers the authors used the phrase "ego close" to designate the ego stance which is referred to here as "reality close." The term ego close proved misleading to many readers and was dropped from this study.

stopped moving, he was to make a heavy dot but leave his pencil on the paper ready to resume tracing without delay if the light started moving again. Should the pencil reach the edge of the board, he was to start again from the approximate center of the sheet and proceed as before. The room was not darkened when these instructions were given, and discerning *Ss* could easily tell that the light was actually a stationary one. No suggestion was given that the light would move, *Ss* were merely told it "may or may not appear to move." The score used in this study for representing autokinesis was the logarithmic conversion of the length of the line drawn by *S* to record the apparent motion of the light.

No pretest period of dark adaptation was provided. The 10-min. test session gave all *Ss* ample opportunity to adapt to the dark. Those who were most prone to see autokinetic movement did not need any time at all to accommodate themselves to these stimulus conditions; they saw autokinetic movement almost from the first moment the room lights were turned off. Those not inclined to see autokinetic movement saw little or none even after full dark adaptation had been achieved.

In a pilot study and two replication studies, inpatient, outpatient, and normal volunteers were given the autokinetic test and then interviewed for their subjective experiences in the autokinetic situation. An interview guide was prepared to help *Ss* describe observations bearing on the two predictions which were to be tested. To learn whether, and to what extent, the objective, three-dimensional space of the experimental room faded out, each *S* was asked a series of progressively more pointed questions, such as the following:

(a) Can you remember your impressions of the room as you were watching the light? Did the room seem any different to you in the dark than it was when you first walked in? (What were your impressions as to its size? Did it seem as if you were in some different kind of room? If another kind of room, what kind?) (b) To what extent did you remain aware of some of the things in the room? Where you were sitting in the room? Where the light was located in the room? (c) When (If) the room seemed so different to you, to what extent did it continue to seem to be a room at all? Did it still feel as if you were indoors?

To assess the degree to which images, personalized phantasies, and preoccupations were projected onto the autokinetic light, each *S* was asked such questions as the following:

(a) What were your impressions of the light? (b) Did the light make you think of anything? Did it ever seem like something else, something other than a point of light? (c) How much like a _____ was it? Was this a passing impression, or did it continue to seem that way throughout the ten minutes? Did the light ever resemble anything else? (d) Did you have any other thoughts about the moving light? Any speculations

about what the course of movement might tell about you? (e) When it seemed like you were watching a ———, where did you seem to be watching it from?

The answers to each set of questions were tape-recorded, transcribed, and excerpted. Excerpts describing reality closeness were pooled for each *S* and separated from those which described phantasy embellishments of the autokinetic experience. Five-point equal interval scales were used and each *S* was given independent ratings for reality closeness and phantasy embellishment.

Ss who scored 1 on reality closeness typically reported continued awareness of the actual size and shape of the room throughout the 10 min. of darkness. In fact, they were puzzled by questions about how the room may have "seemed to change." They remained more or less conscious of objects in the room, of incidental sounds which easily caught their attention, etc. A 5 rating for reality closeness was made when an *S* reported the feeling that he lost all sense of the room he was in during the 10 min. of darkness and that he felt at times as if he were outdoors on a dark night, or out in unbounded space. Implicit in this rating is a distinction between "reality testing" and the "sense of reality." A high rating for reality closeness concerns itself only with changes in the sense of reality; that is, Ss described how the room *felt* to them in contrast to what they knew the room to be like in actuality.

A 5 score for phantasy embellishment was given whenever the light seemed to take on some special meaning or when a phantasy replaced the more objectively neutral light experience, for example, "like watching headlights of an automobile which is driving along a distant mountain road." A 1 score on phantasy embellishment was given for reports in which the light was experienced as being nothing more than a pinpoint light.

Examples of excerpts rated 1 through 5 on reality closeness and phantasy embellishments are available on request.

In the pilot study, 48 inpatient and outpatient volunteers were tested by the first author, who then excerpted those statements about reality closeness and phantasy for independent analysis by three

raters. In the first replication study, 31 patient and normal volunteers were tested and interviewed by a third-year graduate student, and their protocols excerpted by this student for "blind" rating by the authors. In the second replication study, a still more exacting set of experimental conditions was employed. To determine to what extent the relationship the authors were demonstrating was a function of the skill and subtlety of the interviewer, another group of volunteer Ss was tested, but this time by a person with no formal training in psychology. She was an alert, intelligent woman who had been the project secretary for some years. She did the interviewing of 60 Ss and excerpted the relevant passages for the two senior raters.

All interviewers took precautions to avoid seeing their Ss' autokinetic tracings to prevent unconsciously influencing Ss' remarks in the interview which immediately followed the autokinetic test, and, of course, to avoid influencing their ratings for reality closeness and phantasy embellishment. Interrater reliability in scoring the two independent variables was satisfactory. The median reliability coefficient for phantasy was .78 for all raters and .85 for the two senior raters. The median reliability coefficient for reality closeness was .68 for all raters and .77 for the senior raters.

RESULTS

First Prediction: The Relationship of Autokinesis to Reality Closeness

The pilot study and both replication studies showed that it is, in fact, true that the more oblivious one becomes to his concrete surroundings in the autokinetic test situation, the more autokinetic movement he is likely to see (Table 1). In the pilot group of 48 Ss, the correlations between autokinesis and reality closeness ranged from .53 when rated by the least experienced student rater to .77 when rated by one of the authors. All correlations were significant at far better than the $p < .001$ level of confidence.

TABLE 1
RELATIONSHIP BETWEEN AUTOKINESIS AND REALITY CLOSNESS

Pilot study (<i>N</i> = 48)		First replication (<i>N</i> = 31)		Second replication (<i>N</i> = 60)		All cases combined (<i>N</i> = 139)	
Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>
First author (MM)	.77**	MM ₁	.56**	MM	.41**	MM	.57**
Second author (HV)	.66**	MM ₂ ^b	.59**	HV	.46**	HV	.53**
Student rater PL ^a	.59**	HV	.59**				
Student rater MK ^a	.53**	MK	.50*				

^a *N* = 45.

^b Rerated by MM a year later to determine repeat reliability of reality-closeness ratings. Repeat reliability without Yates correction was .79.

* $p < .01$.

** $p < .001$.

TABLE 2
RELATIONSHIP BETWEEN PHANTASY AND REALITY CLOSENESS

Pilot study (<i>N</i> = 44)		First replication (<i>N</i> = 31)		Second replication (<i>N</i> = 57)		All Ss combined (<i>N</i> = 132)	
Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>
First author (MM)	-.38*	MM ₁	-.81***	MM	-.70***	MM	-.47***
Second author (HV)	-.32*	MM ₂ ^a	-.59***	HV	-.41**	HV	-.37***
		HV	-.55**				

^a Rerated by MM a year later to determine repeat reliability of the phantasy ratings. Repeat reliability without Yates correction was .88.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

In the first replication, the 31 Ss who were tested and interviewed by a relatively inexperienced graduate student yielded correlations between autokinesis and reality closeness which ranged .50-.59 ($p < .001$). It is noteworthy that the highest correlations did not come from the rater who had done the interviewing and excerpting but were obtained by the two senior raters who had had no contact with Ss or with any potentially biasing information in the interviews.

In the second replication study ($N = 60$) in which the data were gathered and collated by a relatively unsophisticated, unskilled interviewer, the correlations were more modest (.41 and .46) but still significant at better than the $p < .001$ level. The hypothesized relationship of autokinetic movement to degree of reality closeness was strong enough to overcome measurement errors attributable to interviewer inexperience.

The relationship between autokinesis and reality closeness for the entire group of 139 Ss combined yielded correlation coefficients in the mid-50s for the two raters who had rated all 139 Ss (.53 for one and .57 for the other).

The prediction that autokinesis would be highly correlated with reality closeness in the autokinetic situation was amply supported by these findings.

Second Prediction: The Relationship of Phantasy to Reality Closeness and Autokinesis

The prediction that phantasy emergence will be positively correlated with autokinesis and negatively correlated with reality closeness was similarly confirmed in the pilot study and each of the two replication studies.

Phantasy emergence in the autokinetic experience was closely related to loss of reality anchorage in the test situation (Table 2). The low correlations of the pilot study (-.32 and -.38, $p < .05$) were attributable to occasional insufficiency of interview data about phantasy in the autokinetic experience. The relationship between the two variables became more pronounced when this deficiency was corrected in the replication studies.

In the first replication, the correlation between phantasy and reality closeness ranged from -.55 to -.81 (Table 2). All relationships were in the predicted directions and were significant at, or close to, the $p < .001$ level. In the second replication ($N = 57$), the correlations ranged from -.41 to -.70. The correlations for all Ss combined ($N = 132$), using the more modest MM₂ (first author) data of the first replication study,⁵ were -.47 and -.37, both highly significant.

The relationship of phantasy to autokinesis is shown in Table 3. The more autokinetic movement seen, the more elaborate, vivid, and compelling was the phantasy evoked in the autokinetic situation.

In the pilot study, ($N = 44$), the correlations between autokinesis and phantasy were .44 and .39 when rated by the two senior raters, significant at the $p < .01$ level. Correlations were not statistically significant for the less experienced raters due to a systematic error in their understanding of the criteria which was corrected in the replication studies.

⁵ MM₁ ratings were done by the first author shortly after the first replication interviews were completed and excerpted by MK, one of the student raters. The MM₂ ratings were done by the first author a year later to measure repeat reliability.

TABLE 3
RELATIONSHIP BETWEEN AUTOKINESIS AND PHANTASY

Pilot study (<i>N</i> = 44)		First replication (<i>N</i> = 31)		Second replication (<i>N</i> = 57)		All cases combined (<i>N</i> = 132)	
Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>	Rater	Pearson <i>r</i>
First author (MM)	.44*	MM ₁	.56**	MM	.49**	MM	.51**
Second author (HV)	.39*	MM ₂	.61**	HV	.45**	HV	.48**
Student rater PL ^a	.23	HV	.64**				
Student rater MK ^a	.25	MK	.52*				

^a *N* = 42.

* *p* < .01.

** *p* < .001.

In the first replication study (*N* = 31), correlations range .52-.64, significant at better than the *p* < .001 confidence level. Correlations in the second replication study (*N* = 57) were .49 and .45, both at the same high order of significance. The correlations for all 132 Ss combined were .51 and .48, significant at far better than the *p* < .001 level of confidence.

It should be noted that seven interviews, four in the pilot study and three in the second replication, failed to include questions about phantasy embellishment of the autokinetic experience. Most of these Ss probably would have scored quite low on phantasy judging by the rest of their accounts of the autokinetic experience. An inspection of the data in these cases strongly suggests that, had these few interviews been conducted more carefully, the correlations between phantasy and autokinesis would probably have been even higher than those reported here.

An analysis of these data to see if any of the results reported in the preceding paragraphs were attributable to sex differences or to differences in degree of psychopathology was largely negative. The authors found that all of the relationships reported occurred to a statistically significant extent in men as well as women, but were more pronounced (or at least more easily demonstrated) in women. This may reflect the fact that men sometimes seem more constrained in the autokinetic situation, less inclined to accept, enjoy, indulge themselves in, or perhaps admit to such relaxed, essentially passive states of consciousness.

With respect to the saliency of the predicted relationships in normals as compared with patients, no consistent differences were

found. The relationships were as pronounced in patients as in normal Ss, despite the obvious uneasiness which some patients showed when asked to describe the kinds of subjective experiences with which the study concerned itself.

DISCUSSION

This study was designed with two purposes in mind. One was to determine why the autokinetic test is so effective a measure of reality closeness. The other was to add to the understanding of the autokinetic phenomenon.

In previous studies, the autokinetic score proved to be closely related to a continuum of character structures and life styles. Normal Ss with low-autokinetic scores were typically characterized as outgoing, responsive, other-directed people with perhaps a certain histrionic flair and vivid "presence." Low-autokinetic people who were emotionally disturbed were typically described as emotionally labile, "alloplastic," externalizing, and prone to rely more heavily on the defenses of repression, denial, and acting out. In contrast, people with high-autokinetic scores were more likely to be characterized as detached, intellectualizing, deliberate, and inclined to isolate themselves off from strong feelings and impulses. The more normal high-movement Ss were self-conscious, thoughtful, quiet but stubbornly independent people. The more disturbed high-movement Ss indulged heavily in obsessive rumination and schizoid phantasy. In short, the lower his autokinetic score, the more likely it is that a person lives his life in a close, vivid, direct, interactional involvement with the people and things around him. The higher his autokinetic score, the more likely

it is that a person experiences himself and his world with more detachment and with more subjectivity.

The authors suggested that the common denominator in the life styles, experience styles, and defensive styles of the low autokinetic Ss was their "closeness" to reality, and the common factor in high-autokinetic people at the other end of the continuum was their greater distance or detachment from the compelling "pull" of their surroundings.

What remained unexplained in these studies was why the autokinetic test provided so accurate an index of a person's characteristic position on the reality-closeness-reality-distance continuum. The present study helps fill this gap by showing that the reality closeness by which the authors previously characterized the adaptive styles of low-autokinetic Ss is demonstrable in a very concrete, almost literal sense in the autokinetic situation. Low-autokinetic Ss remain more focally aware of the concrete realities of the experimental room and are much less inclined to drift into phantasy or reverie than are high-autokinetic Ss. The amount of autokinetic movement seen by Ss in the present study is directly related to the degree to which they retained a reality-orienting image of the experimental room throughout most or all of the 10 min. of darkness. This conclusion is arrived at only inferentially here; it is given direct experimental support in another study.⁶ The conclusion draws on, and lends further support to, frame-of-reference explanations of autokinesis reviewed by Royce et al. (1966) and by Reinwald (1952).

The aforementioned formulations draw heavily on psychoanalytic ego psychology in defining qualities of reality contact, forms of engagement and interaction between "ego" and "reality." In Rapaport's (1958) terms, those people who are characteristically more autonomous from reality need more "stimulus nutriment" to keep them in touch with reality. They readily slip into subjective preoccupations, phantasies, or reveries when external conditions are conducive to such a turn of mind. On the other hand, those persons who live their lives in close touch with objects and

events might be said to live with an excess of stimulus nutriment. Much of their awareness of inner psychological events is drowned out by the clamor of external experiences. For them, external reality is so compelling that they find it difficult to detach themselves from external objects and turn attention to subjective experiences.

There are two ways of viewing such individual differences in cognitive style. One might argue that Ss who cling to reality and prevent the emergence of phantasy do so defensively, that is, turn to reality in flight from endopsychic experiences. Similarly, one can hold that those who drift off have learned to escape from the exigencies of reality by withdrawing defensively into themselves. However, there is reason to maintain that, in part, perhaps in large part, individual differences in reality closeness-reality distance are not simply a product of developmentally acquired adaptive or defensive ego styles. They may derive ultimately from constitutionally based ego dispositions (Hartmann, 1939; Rapaport, 1951, 1958). There are some telling data which point in this direction and which urgently call for confirmation. Eysenck and Prell's (1951) study of autokinesis in identical and fraternal twins found that identical twins correlated .73 and fraternal twins only .23 in the extent of autokinetic movement seen. Pending confirmation of these findings by replication studies, Eysenck and Prell's study supports the hypothesis that reality closeness-reality distance may well be a constitutionally based ego disposition.

These formulations about reality closeness and reality distance should not be confused with Witkin's concepts of field dependence-field independence, despite their apparent similarities. On Witkin's continuum, field dependence is a less desirable state than field independence. It indicates a degree of developmental immaturity, of lesser "psychological differentiation," less stable ego identity structures, etc. Reality closeness in the authors' studies carries no such pejorative implications. Low-autokinetic Ss are not necessarily lacking in psychological differentiation. On the contrary, many of the reality-close Ss are highly gifted, intelligent, competent people, with, if anything, a more emphatic sense of identity

⁶ M. Mayman, H. M. Voth, and L. Coyne, manuscript in preparation.

than was found in many reality-distant Ss. The authors predicted, in fact, an orthogonal relationship between field dependence and reality closeness, a prediction which was tested and confirmed recently by Cancro and Voth (1969). One hundred and five paid, normal volunteers were administered Witkin's rod-and-frame and Embedded Figures Tests to determine if these measures correlated with Ss' scores on the autokinetic test. The Es could find no demonstrable relations, linear or curvilinear, between autokinesis and Witkin's measures of field dependence-field independence. This result cannot be dismissed as a measurement artifact, because the rod-and-frame scores correlated highly with the embedded-figures scores, much as they have done in all of Witkin's work with these tests. It seems clear that, operationally as well as conceptually, the Mayman-Voth explanatory constructs and Witkin's are unrelated psychological variables.

REFERENCES

- BUSH, M., HATCHER, R., & MAYMAN, M. Reality attentiveness-inattentiveness and externalization-internalization in defense style. *Journal of Consulting Psychology*, 1969, 33, 343-350.
- CANCRO, R., & VOTH, H. M. Autokinesis and psychological differentiation. *Perceptual and Motor Skills*, 1969, 28, 99-103.
- EYSENCK, H. J., & PRELL, D. B. The inheritance of neuroticism: An experimental study. *Journal of Mental Science*, 1951, 408, 339-465.
- HARTMANN, H. *Ego psychology and the problem of adaptation*. New York: International Universities Press, 1939.
- RAPAPORT, D. The autonomy of the ego. *Bulletin of the Menninger Clinic*, 1951, 15, 113-123.
- RAPAPORT, D. The theory of ego autonomy: A generalization. *Bulletin of the Menninger Clinic*, 1958, 22, 13-35.
- REINWALD, F. L. An experimental investigation of conditions affecting autokinetic sensations. Unpublished doctoral dissertation, University of Texas, 1952.
- ROYCE, J. R., CARRAN, A. B., AFTANAS, M., LEHMAN, R. S., & BLUMENTHAL, A. The autokinetic phenomenon: A critical review. *Psychological Bulletin*, 1966, 65, 243-260.
- SANTOS, J. F., FARROW, B. J., & HAINES, J. R. Consistency and style of autokinetic movement. *Perceptual and Motor Skills*, 1965, 21, 583-586.
- SCHWARTZ, M., & SHAGASS, C. Response to colored and conflict-inducing stimuli in a psychiatric population. *Perceptual and Motor Skills*, 1960, 11, 245-252.
- TEMERLIN, M. K. One determinant of the capacity to free-associate in psychotherapy. *Journal of Abnormal and Social Psychology*, 1956, 53, 16-18.
- VOTH, A. C. Individual differences in the autokinetic phenomenon. *Journal of Experimental Psychology*, 1941, 29, 306-322.
- VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *American Journal of Psychology*, 1947, 103, 793-805.
- VOTH, H. M. Choice of illness. *Archives of General Psychiatry*, 1962, 6, 149-156.
- VOTH, H. M., & MAYMAN, M. A dimension of personality organization. *Archives of General Psychiatry*, 1963, 8, 366-380.
- VOTH, H. M., & MAYMAN, H. The psychotherapy process: Its relation to ego-closeness-ego-distance: Part I. *Journal of Nervous and Mental Disease*, 1966, 143, 324-337.
- VOTH, H. M., & MAYMAN, M. Diagnostic and treatment implications of ego-closeness-ego-distance: Autokinesis as a diagnostic instrument. *Comprehensive Psychiatry*, 1967, 8, 203-216.
- VOTH, H. M., & MAYMAN, M. The psychotherapy process and its relation to ego-closeness-ego-distance. II. *Journal of Nervous and Mental Disease*, 1968, 147, 308-315.

(Received October 13, 1967)

SELF-CONCEPT THERAPY FOR ADOLESCENT FEMALES¹

C. W. COLE,² E. R. OETTING, AND R. W. MISKIMINS

Colorado State University

Fourteen adolescent females referred for delinquent and acting-out behaviors were compared pre- and posttreatment on measures of self-concept, ideal self-concept, and anxiety. These Ss were exposed to a 10-wk. socialization program designed to produce positive changes in self-concept; however, 9 Ss were in a group run by community volunteers while 5 Ss were in a group run by mental health professionals. Self-concept changes in a nonreferred comparison group of 8 nondelinquent Ss appeared nearly random in contrast to the systematic improvement for the 14 Ss in therapy. There was some evidence of differential affectiveness in that Ss in the professionally led group appeared to undergo the most marked changes.

There is considerable support for the contention that adolescents who demonstrate difficulty in school coupled with "delinquent" nonschool behavior (sexual acting out, stealing, etc.) are relatively easily described through the use of a "self-concept" theoretical framework. Many investigators have found that delinquent or "behaviorally disturbed" adolescents have a self-concept quite different from that of nondelinquent children (e.g., Curtis, 1963; Reckless, Dinitz, & Day, 1957). Implicit in such studies is the suggestion that treatment for behaviorally deviant adolescents at self-concept alteration; however, there has been no clear explication of a therapeutic mechanism for producing such change other than the traditional approach through long-term individual therapy. This study reports the results of a relatively restricted type of group treatment program aimed specifically at producing self-concept changes in female adolescents with a record of behavior problems.

METHOD

Subjects

Of an original pool of 38 adolescent females, 22 Ss were compared pre- and posttreatment on measures

¹ This investigation was supported, in part, by a Public Health Service Research Grant R 11-MH-00458 (04) from the National Institutes of Health, Public Health Service. A more detailed report of the self-concept program, including brief case studies, may be obtained by writing the Executive Director, Southeast Wyoming Mental Health Center, Cheyenne, Wyoming. Write for Technical Report VII, "Community Resources in Treatment."

² Requests for reprints should be sent to Charles W. Cole, Department of Psychology, Colorado State University, Fort Collins, Colorado 80521.

of self-concept and anxiety. These Ss ranged in age 13-17 and were located in three comparison groups: 5 Ss (Group P) were in a group therapy experience run by professional mental health workers from the Southeast Wyoming Mental Health Center, Cheyenne, Wyoming; 9 Ss (Group Q) were in a group of similar intent and design which was run by the Cheyenne chapter of Quota International, a volunteer service organization for women; and 8 Ss (Group N) were selected from the public schools to provide a "normal" comparison group. The Ss in both the professional and volunteer groups were adolescent females who had been referred by various treatment and welfare agencies as delinquent or as having less severe behavior problems. All were considered by school authorities as potential drop-outs. A large number of these girls had been identified as acting out sexually. The Ss in the nontherapy comparison group were similar in age and academic performance, but had no record of acting-out behavior. This group was used to provide normative data as well as to provide a control for changes which might occur simply as a function of time.

Instrumentation

Pre- and posttesting for the three groups was performed over a 10-wk. interval. The measures used were identical at both testings. The test battery provided measures of the concepts "Me as I am today" and "My ideal-self." An index of the evaluative attitude of each S toward these concepts was provided by a scale of 20 adjective pairs presented in a semantic differential format. All items had been selected for their high loadings on the evaluative factor of meaning as indicated by the work of Osgood, Suci, and Tannenbaum (1957). Representative pairs are good-bad, valuable-worthless, etc. Summing down these 20 scales yielded total "value" score for each concept. Each of the concepts was rated against the same set of scales, thus allowing the computation of a score representing a self-ideal discrepancy along the value dimension, a commonly used measure of self-regard. The Anxiety Differential (Alexander & Husek, 1962; Husek & Alexander,

1963) was used to provide a measure of situational anxiety.

Procedure

The two "experimental" therapy groups were put through a 10-wk. program designed to produce positive self-image changes in the areas of personal and social appearance. The program was presented, not as "therapy" but as a program concerned with training in socially relevant skills. Eight major topics dominated the sessions: make-up, hair styling, exercises and diet, clothes and accessories, posture and general demeanor, dating etiquette, review of make-up and hair styling, and problems in dining out. Walking, dancing, the art of conversation, and ease in social situation were stressed during the weekly sessions. Volunteers in the community spoke to the group on many of these topics. The speakers were selected for their value as feminine identification figures. One, for example, was a local beauty-shop operator. In addition, both of the group leaders could be described as mature, relaxed, and actively feminine.

The program was planned to offer three direct approaches to the modification of the self-concept. First, the individuals selected to run the groups provided appropriate feminine role models. It was intended that these leaders would establish relationships with the girls which would permit the girls to incorporate some of the characteristics of the leaders into their own self-concepts. Second, the content of the program focused directly on the appearance and the social behavior of the girls. Much of the program could be conceptualized as straightforward "attractiveness training." The basic procedure used for shaping new attitudes and behaviors in the girls was to change their appearance (e.g., giving them a free beauty treatment) or, in the case of a particular kind of behavior, to have them "role play" the behavior. Positive changes in either appearance or behavior were then reinforced through the social approval of the leaders and the other members of the group. Finally, in order to create a situation in which "new ways of being me" could be tried, group discussions were focused on the acceptance of personal limitations, and attempts were made to develop an "in group" feeling in order to give the girls a sense of being accepted by their peers.

It was assumed that the provision of appropriate feminine role models accompanied by the teaching of feminine role behaviors would provide a milieu in which positive alterations in the self-concepts of Ss would occur and that, by making the members of the group feel more personally attractive, they would be less likely to seek to prove their femininity through socially unacceptable behavior. It is, perhaps, apparent that such a social learning approach is predicated on a behavioristic conception of the origin of affective states. In this case, it would be held that the self- and related concepts are formed through interactive experiences with others. While such states may mediate subsequent behaviors, they

are amenable to change through the provision of a new environment and a corrective set of interpersonal interactions.

RESULTS

The mean self-ideal self-concept discrepancies of all Ss were arithmetically transformed to eliminate negative values. The mean pretest self-ideal discrepancies for the nontreated group of normal Ss and for the professional and volunteer-led groups were 37.50, 28.00, and 36.20, respectively. While the means of the three groups did not differ significantly, the variances obtained were quite dissimilar (Group N, $s^2 = 54.50$; Group P, $s^2 = 386.56$; Group Q, $s^2 = 449.25$). The variances of each of the two therapy groups were significantly greater than the variance of the normal group (Group P, $F = 7.09$, $p < .05$; Group Q, $F = 8.24$, $p < .01$); however, the two therapy groups did not differ significantly ($F = 1.16$). While the pretest anxiety scores correlated .473 ($p < .005$) with the self-ideal discrepancy scores when all three groups were combined, there were no significant differences between the therapy and normal comparison groups.

At the time of posttesting, there were still no significant mean differences between the three groups; however, contrary to the pretest findings, the variances of the groups no longer differed significantly (Group N, $s^2 = 74.36$; Group P, $s^2 = 27.20$; Group Q, $s^2 = 150.91$). While the control group retained approximately the same range, there were less extreme scores in the treatment groups, particularly in Group P. Again, there were no significant differences in measured anxiety. In all groups, a tendency for more positive self-evaluation was observed.

Although there were no significant differences among the variances of the three groups at the time of posttesting, there were differing levels of change over the 10-wk. interval. Over the pretest to posttest period, the non-behavior-problem Ss changed very little, while the variance of self-ideal discrepancy in the volunteer-led therapy group declined from 449.21 to 150.91 ($F = 2.98$). The reduction in variance for the professionally led therapy group was greater—changing from 386.56 to 27.20 ($F = 14.21$, $p < .001$). Sixty-

two percent of the Group N Ss moved to less extreme self-ideal discrepancy scores, while 38% obtained more extreme scores. In the combined therapy groups, however, 93% of the Ss had less extreme scores on posttesting, with only 7% tending toward more extreme self-ideal discrepancy. This movement was accounted for primarily by changes in the level of value placed on the self-concept rather than by ideal self changes. In general, extremely high self-concepts were lowered, while extremely low self-concepts were increased.

DISCUSSION

There are several investigations which suggest that the relationship between adjustment and self-regard is basically curvilinear (e.g., Block & Thomas, 1955; Chordorkoff, 1954; Miskimins, 1967). These studies indicate that maladjustment may be reflected either in very high or very low self-ideal discrepancies. Under this assumption, it may be stated that both therapy groups (P and Q) were, at the time of pretesting, more maladjusted than the comparison group of nonreferred normals. While many of the behaviorally disturbed adolescent girls evidenced high self-ideal discrepancy, there were some who showed extremely low incongruence even to the extreme of rating the self higher than the ideal self-concept. Block and Thomas (1955) and Zuckerman and Monashkin (1957) have proposed that the "high self-acceptor" may in fact be more severely disturbed than the individual for whom a high self-ideal discrepancy may be obtained. This is consistent with the finding of Suinn and Hill (1964), replicated by the present study, to the effect that anxiety and self-acceptance are negatively related. In the therapy groups, many of the girls were defined by their test data as extremely defensive; none of the nonreferred comparison Ss fell into this classification. The usual interpretation is that extremely high self-acceptors are defensive individuals who find their behavioral tactics useful in minimizing feelings of anxiety, increasing day-to-day personal comfort, and in avoiding the distresses of negative self-regard.

It should be noted that there were no extremely small self-ideal discrepancies among

treated Ss at the time of the posttest. Perhaps the self-concept program was particularly effective in dealing with adolescent females with this type of problem. It would appear that these girls became much more self-reflective, open, and spontaneous. Because of the nonspecific nature of the measure used, some evidence is provided in support of the contention that these changes in self-concept were not confined simply to the areas dealt with in the groups (e.g., grooming and hygiene)—the changes which took place were very likely global. If so, generalization to behaviors in the extratherapy setting should have been enhanced.

The instrument used in assessing the self- and ideal self-concepts was not designed to yield a general self-description, but rather to provide a measure of the level of value which each S associated with either of these concepts. Therefore, it is difficult to say whether Ss in the therapy groups saw themselves or their ideal selves as having different characteristics than did the comparison group of nondelinquent females. A similar score would mean, however, that they attached the same level of value to the concept.

Perhaps the level of self-value and the self-ideal discrepancy are critical to adjustive behavior. For at least the 6-mo. period following the self-concept program, none of the girls was reported to be in serious trouble; none had dropped out of school for either disciplinary or academic reasons.

It cannot be said with any certainty that the critical factor in this apparent behavioral improvement was the "normalization" of the self-ideal-self relationship, but whatever the case, it may be concluded that there is considerable evidence for the efficacy of both professional and volunteer-led programs in producing positive changes in the self-concepts of behaviorally disturbed adolescent females. The Ss in the self-concept program provided, at termination, levels of self-regard more typical of normal adolescents, while changes in the nonproblem comparison group appeared to be nearly random. Furthermore, this change was accomplished by a direct program concentrating on the learning of personal and social skills. It is also likely that the opportunity for constant interaction with highly adequate

feminine role models was a major ingredient in the program's success. An advantage of such a program is its high face validity. The girls did not feel that they were in "treatment"; yet, there is strong evidence that they received many of the benefits usually associated with traditional therapeutic approaches.

While it is apparent that both of the therapy groups were effective to some degree, there is a question as to the differential effectiveness of the two. There is some evidence that more marked self-concept changes were produced in the group led by mental health professionals. It is possible that, although the two programs were ostensibly identical, the volunteers may differ in basic attitudes about what they are doing. There is some evidence, for example, that they "lectured" more, and "discussed" less. Whatever the reason, the practical aspects of the self-concept program should be considered. Trained professionals are harder to obtain for work such as this and, of course, must be paid, whereas the use of community volunteers provides a much larger pool of persons to conduct self-concept groups along with a considerable reduction in expense. Furthermore, certain adolescents may be equally benefitted by either professional or volunteer-led groups. A selection process which would provide a pretreatment rationale for group assignment seems highly desirable as a way of providing for the most effective use of available mental health personnel.

REFERENCES

- ALEXANDER, S., & HUSEK, T. R. The anxiety differential: Initial steps in the development of a measure of situational anxiety. *Educational and Psychological Measurement*, 1962, 22, 325-348.
- BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *Journal of Abnormal and Social Psychology*, 1955, 51, 254-259.
- CHODORKOFF, B. Adjustment and the discrepancy between the perceived and ideal self. *Journal of Clinical Psychology*, 1954, 10, 266-268.
- CURTIS, G. R. Self-concept and identification in pre-schizophrenic adolescents. Unpublished doctoral dissertation, University of Southern California, 1963.
- HUSEK, T. R., & ALEXANDER, S. The effectiveness of the anxiety differential in examination stress situations. *Educational and Psychological Measurement*, 1963, 23, 309-318.
- MISKIMINS, R. W. Self concept and psychopathology: A review. Unpublished manuscript, University of Colorado, 1967.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- RECKLESS, W. C., DINITZ, S., & DAY, B. The self component in potential delinquency and potential non-delinquency. *American Sociological Review*, 1957, 22, 566-570.
- SUINN, R. M., & HILL, H. Influence of anxiety on the relationship between self-acceptance and acceptance of others. *Journal of Consulting Psychology*, 1964, 28, 116-119.
- TAYLOR, D. M. Changes in the self concept without psychotherapy. *Journal of Consulting Psychology*, 1955, 19, 205-209.
- ZUCKERMAN, M., & MONASHKIN, I. Self-acceptance and psychopathology. *Journal of Consulting Psychology*, 1957, 21, 145-148.

(Received March 10, 1967)

SUBJECTIVE MATCHING OF ANXIETY TO INTENSITIES OF WHITE NOISE

RICHARD SULLIVAN¹

New York University Medical Center

Twenty Ss were asked to match intensities of white noise to levels of experienced anxiety under stress and nonstress conditions. Anxiety measures were obtained through the use of a rating scale. The Ss were also asked to judge a minimal aversion threshold (MAT) of white noise and to rate a level of anxiety which matched the aversiveness of the noise. Anxiety and noise matchings indicated relatively lawful agreement during stress and nonstress conditions. It was found that at only one point during the stress condition did the anxiety and noise judgments exceed the mean subjective aversion level. The findings suggest that changes in the intensity of internal states may be measured through the use of judged matchings with changes in intensive sensory stimuli.

Stevens' (1959) validation of the psychophysical "power law" through use of comparative cross-modal judgments demonstrated that Ss are capable of lawfully relating the changes in one sensory mode to that of a second sensory mode. Stevens (1961) has also shown that "direct" methods of psychological scaling may be used to measure the subjective experience of changes in the magnitude of stimuli which have no related physical metric. The requirements of such scaling are that the stimulus continuum varies in experienced magnitude and can be assigned an ordinal rank by *E* according to a specified criterion.

The success of Stevens' techniques for cross-modal matching suggests that the subjective experience of anxiety may be measured through use of similar procedures. Anxiety may be considered an intensive continuum wherein experienced amplitudes vary relative to the stress of differing conditions. Initially, criteria for the assignment of ranks to different conditions of stress may be based on the S's rated judgment of those conditions. However, before attempting to have Ss directly estimate their levels of anxiety under different conditions it would have to be established that they could judge anxiety as an intensive continuum. A cross-modal comparison of intensities of anxiety with that of

another intensive continuum seemed to be an appropriate means of obtaining that information. The present study was designed to determine whether Ss could match intensities of white noise to levels of experienced anxiety under different conditions of stress.

A second aspect of the study dealt with Ss' ability to judge a minimal aversion threshold (MAT) of noise and to estimate a level of anxiety which would match the aversiveness of the noise. These response measures were obtained on the assumption that since many intensive sensory continua become aversive at high subjective amplitudes, an origin or MAT could be specified above which all further increments in amplitude of stimulation would be considered aversive. Thus, if anxiety were to be meaningfully related to noise levels, when the rated anxiety was above the MAT level, the matching noise judgment should also be above the MAT established for noise. The possibility that a factor of sensory aversiveness may alter judgment (Ekman, Frankenhauser, Levander, & Mellis, 1964; Stevens, Carton, & Shickman, 1958; Sullivan, 1968a, 1968b) is similar to the assumption that anxiety, at high levels of subjective magnitude, may significantly influence subjective judgment.

METHOD

Subjects. Six female and 14 male students (ages 19-52) in an introductory psychology course served as Ss.

Apparatus. The apparatus consisted of a white noise generator in circuit with a key and an 8-ohm speaker. The speaker was placed in the center of a

¹ Also at the City College of the City University of New York.

Requests for reprints should be sent to the author, New York University Medical Center, 550 First Avenue, New York, New York 10016.

semicircle of Ss seated in an ordinary classroom. The ambient noise level of the room varied 57–62 db. over various testing periods. White noise was presented to Ss in all conditions over a continuous 2-min. interval for both ascending and descending trials. Sound pressure levels (SPL) at various positions in the semicircle for both ambient and experimental noise was measured by a General Radio sound level meter. Differences in sound levels within the room (maximum was 6 db. at distal sides) were accounted for by adjusting a given S's judgment in terms of his position. The dial settings on the noise generator were arbitrary figures used to reduce E's possible bias in changing the noise levels over the 2-min. interval of presentation.

The anxiety rating scale consisted of a sheet of paper with a 7-in. line at the top having intervals spaced 0–7. Below the line, definitions of the intervals were given as follows: 0–1, no anxiety, perfectly relaxed, and comfortable; 1–2, little anxiety, feel comfortable, and pleasant; 2–3, some anxiety is present but not bothersome; 3–4, anxiety is present, not intense or bothersome; 4–5, anxiety is moderately intense and somewhat uncomfortable; 5–6, anxiety is fairly intense and unpleasant but not intolerable; 6–7, anxiety is intense, unpleasant, and intolerable. The Ss in all conditions were asked to place an "X" on the line in a position which best described the anxiety they were experiencing at the time of rating.

Procedure. In general, anxiety ratings and judgments of minimal aversion thresholds, and noise matchings were spread over various lecture hours (nonstress condition). Brevity in both ratings and noise judgments was crucial since Ss were to be asked to do both at five time intervals during the midterm examination (stress condition). The first half of the nonstress anxiety ratings, matched noise judgments, and minimal aversion judgments of anxiety and noise were obtained prior to the midterm exam; the second half of judgments and ratings was obtained between the midterm and final exam period. During the final exam, Ss were asked only to rate their levels of anxiety and were not asked to match those levels to intensities of white noise.

MAT: Noise judgments. The Ss were asked to judge the level of white noise which they considered to be minimally aversive. The minimal aversion level of noise was defined as that intensity of noise which is at the point of becoming annoying or bothersome. The procedure for obtaining responses was to have all Ss close their eyes on signal and have them raise their hands when they felt the noise to be minimally aversive. An assistant of E's took down each S's judgment by recording the dial setting on the noise generator when S raised his hand. Two trials were given at each of the four sessions. A trial consisted of a presentation of the noise at 50 db. (SPL re: .0002 dynes/cm²) which continuously increased over a 2-min. interval to an upper limit of 95 db.²

² Since no S went above 92 db. in judging the MAT, it was felt that 95 db. would be a suitable upper limit for the matching task. The possibility that this upper limit might have influenced judgments is discussed later.

Following a 30-sec. pause, the noise was presented at 95 db. and decreased over a 2-min. interval to the lower limit of 50 db. A 2-min. rest period was given between the two trials. During the descending series, Ss were asked to raise their hands when they felt that the noise was no longer bothersome, annoying, or minimally aversive.

MAT: Anxiety ratings. On four separate occasions, Ss were asked to judge their minimal aversion threshold for noise by rating the anxiety scale in terms of what level of anxiety they felt would be the equivalent of the MAT noise.

Noise and anxiety matching: Nonstress condition. The Ss were requested to rate themselves on the anxiety scale at five different times during the four lecture periods (5, 30, 55, 80, and 105 min., respectively). After each rating, the sheets were collected and Ss were then asked to judge the intensity of white noise which matched the level of anxiety they were experiencing at that time. One ascending series and one descending series was given for each anxiety rating.

Noise and anxiety matching: Stress condition. During the midterm exam period, Ss were asked at the five different times to rate their level of anxiety. Following the collection of the rating sheets, they judged the white noise intensity which matched their level of anxiety. The first rating and judgment preceded the beginning of the exam (5-min. judgment); the fifth rating and judgment followed the completion of the exam (105-min. judgment). The procedure for obtaining responses was the same as outlined in the preceding paragraph. On a signal from E all Ss would pause, rate their anxiety, pass their sheets to E or E's assistant, close their eyes, and judge the noise. Two trials (one ascending and one descending series) were given for each anxiety rating. The Ss were now fairly practiced in the procedure and the ratings and judgments at any given interval took approximately 7 min.

RESULTS

Anxiety ratings and noise judgments obtained under the various conditions are plotted in Figure 1 (anxiety) and Figure 2 (noise). In both figures, the mean level of anxiety or noise which was judged to be minimally aversive in the nonstress condition is represented by a straight line (visual best fit) in the upper portion of the graph. (Mean MAT ratings for anxiety were 4.34, 4.55, 4.28, and 4.40; mean MAT judgments of noise were 85.18, 85.36, 84.70, and 85.53 db.).³ As seen in Figure 1, mean ratings of anxiety experienced at the five different times in the nonstress condition were relatively stable (lower curve). Mean

³ There were no appreciable differences between median and mean measures in both the anxiety ratings and noise judgments.

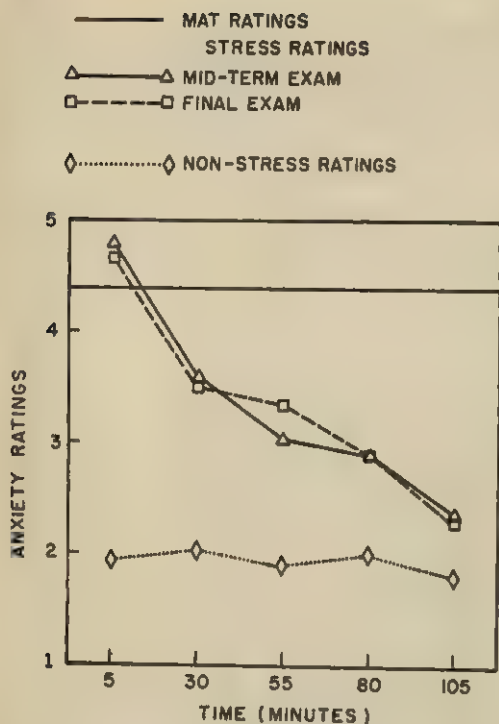


FIG. 1. Mean anxiety ratings at five time intervals during stress (exams) and nonstress conditions (lecture periods). (Straight line in the upper portion of the graph represents the mean rating of minimal aversiveness of anxiety.)

levels of rated anxiety during both stress conditions (midterm and final exam periods) were essentially the same for each of the time intervals. Anxiety was rated highest prior to the exam (5-min. interval) and lowest at the completion of the exam (105-min. interval). A t test for matched pairs at the 105-min. time interval between stress and nonstress ratings was found to be significant ($t = 3.59$, $df = 19$, $p < .005$).

Similar to the anxiety ratings in the non-stress condition, the judgment of noise intensities matched to levels of anxiety at various times during the nonstress condition indicated relative stability over the 105-min. period (Figure 2, lower curve). During the stress condition, the intensity of noise matched to the level of anxiety was highest prior to the exam (5-min. judgment) and lowest at the completion of the exam. (105-min. judgment). The only point at which the noise judgments and anxiety ratings were above the

MAT was prior to the exam at the 5-min. time interval.

Table 1 presents a four-way analysis of variance of noise judgments matched to anxiety levels at different times under stress and nonstress conditions (stress versus nonstress, time intervals, ascending versus descending series, S_s). The main effect of stress was found to be significant ($F = 35.89$, $df = 1/19$, $p < .001$) as was the main effect of series (ascending versus descending trials, $F = 178.49$, $df = 1/19$, $p < .001$). The significant Stress \times Time interaction ($F = 6.90$, $df = 4/76$, $p < .001$) indicated that, in terms of noise matching with anxiety, different levels of anxiety were experienced at the different time intervals. The significant Stress \times Series \times S_s interaction indicated, as might be expected, that different S_s equated the noise intensities to different levels of anxiety at the differing times of judgment. A trend analysis of the noise matchings was done following the procedure outlined in Edwards (1960) and indicated that the linearity of the slope of judg-

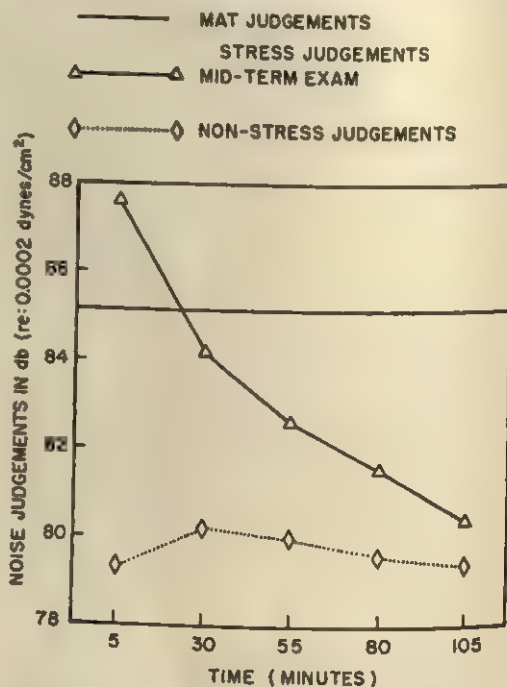


FIG. 2. Mean noise judgments at five time intervals during stress (exam) and nonstress conditions (lecture periods). (Straight line in upper portion of the graph represents the mean judgment of minimal aversiveness of white noise.)

ments was significant ($F = 27.9$, $df = 1/60$, $p < .001$).

A significant finding in the present study was that descending noise judgments were higher than ascending noise judgments regardless of the task (MAT judgment or anxiety matching). Means of the ascending and descending judgments of noise matchings at the five time intervals during the stress condition are presented in Table 2.

The significant Stress \times Time \times Series interaction ($F = 32.48$, $df = 4/76$, $p < .001$) suggests that judgment variance within the ascending and descending trials may have altered depending on the level of anxiety experienced by Ss at the time of judgment.

DISCUSSION

The findings indicate that Ss were capable of matching changes in the level of experienced anxiety with changes in amplitude of white noise. The results further indicate that when anxiety was judged to be above a rated minimal aversion level, the matching noise level was also above the minimal aversion threshold for white noise. The findings suggest, therefore, that a reasonably lawful measurement of anxiety may be obtained through use of cross-modal matchings of anxiety with other intensive continua.

The mean level of white noise judged to be minimally aversive (85 db.) is approximately the same as that reported by Spieth (1956) for annoyance thresholds of Ss who imagined a noisy work environment while making judgments. The similarity in findings may have been due to the relatively high level of

TABLE 2

MEAN ASCENDING AND DESCENDING NOISE JUDGMENTS AT FIVE INTERVALS IN THE STRESS CONDITION

Trial	Time interval (in min.)				
	5	30	55	80	105
Ascending (db.)	86.5	82.5	81.1	79.9	78.3
Descending (db.)	89.2	86.1	84.4	83.8	82.0

ambient noise (57–62 db.) in the classroom while judgments were made by the present Ss. Other factors which may have contributed to the 85-db. MAT were adaptation changes occurring due to the continuous presentation of noise and anchoring effects. In a related study of the effect of adaptation on MAT noise judgments, Sullivan, Warren, and Dabice used adaptation levels of 70, 85, 95, and 110 db. Their findings indicated that the series effect (ascending versus descending trials) had a greater influence on noise judgments than the adaptation effect. The difference between ascending and descending judgments in the present study may be due to Ss having made relative judgments on the basis of the upper and lower anchoring noise levels. Anchoring effects seem also to have influenced anxiety ratings. Thus, while most Ss reported great relief from anxiety at the end of the stress period, their anxiety ratings were still significantly higher ($t = 3.59$, $df = 19$, $p < .005$) than the ratings given at the 105-min. interval in the nonstress period.

The procedure of instruction and practice appears to have had an important influence on the present results since most Ss expressed a preference for judging noise tolerance rather than noise matching. This preference seems to reflect the possibility of a summative effect of noise and anxiety in interfering with task performance. The presentation of an aversive noise may make an anxiety level appear subjectively more intense or an aversive level of anxiety may make a noise appear more intense. In a review of the effects of noise on behavior, Broadbent (1957) suggested that high anxious individuals have a low tolerance for noise. The implication of Broadbent's suggestion is that judgments of the aversiveness of noise stimuli are dependent on the level of

TABLE 1

FOUR-WAY ANALYSIS OF VARIANCE OF NOISE JUDGMENTS MATCHED TO LEVELS OF ANXIETY

Source	df	MS	F
Stress (A)	1	1270.5	35.89*
Time (B)	4	197.0	13.97*
Series (C)	1	1463.6	178.49*
A \times B	4	161.5	6.90*
A \times C	1	109.5	43.80*
B \times C	4	16.8	1.56
A \times B \times C	4	68.2	32.48*
A \times B \times C \times D	76	2.1	

Note.—Stress versus nonstress; five time intervals in ascending and descending series. Components of the variance analysis for three fixed and one random variable was done according to procedure outlined by Hicks (1964).

* $p < .001$.

an S's anxiety. However, other variables such as previous experience with different noise environments (Spieth, 1956), the individual's subjective definition of aversiveness (Wolff, 1964), and whether the noise is intrinsic or extrinsic to the task would all appear to influence the relationship between noise tolerance and anxiety. Using either the anxiety ratings or the matching noise judgments of minimal aversion as a criterion, it appears that task involvement reduced the level of anxiety experienced by Ss (5- versus 30-min. judgments in the stress condition). Task involvement also appeared to influence judgments of the annoyance of noise stimuli (Kryter, 1966).

Although the distributions of anxiety ratings appeared to be normal, no inferences can be made with regard to the validity of the scale. A more appropriate measure of anxiety ratings may have been obtained through use of the method of successive interval scaling (Guilford, 1954; Torgerson, 1958). Changes in the rating styles during stress and nonstress conditions could then have been analyzed. However, under conditions where the subjective state alters considerably over brief periods of time, direct methods of estimation would seem to be a more practical means of measurement. There is evidence which suggests a lawful relationship between interval and ratio forms of scaling procedures (Ekman & Künnapas, 1962; McGill, 1960). Until that relationship is shown to be valid, Stevens' (1961) criticism of the use of category scales to measure intensive continua would suggest that such measurements may not apply to anxiety. Thus, the use of direct methods to measure anxiety may be a more appropriate and practical procedure than the use of category-type scaling techniques. However, even the direct methods of measurement may not apply when stimulus intensities are within an aversive range of stimulation. It would, therefore, have to be determined that judgments of aversive stimuli have dispersion characteristics similar to those of nonaversive stimulus ranges.

Whether or not anxiety can be directly related to magnitude changes in intensive sensory continua, it does appear possible to monitor changes in anxiety states by use of

judged variations in matchings to intensive continua. The same procedure would seem to apply to other subjective states, such as clinical pain, hunger, thirst, and fatigue.

REFERENCES

- BROADBENT, D. E. Effects of noise on behavior. In C. M. Harris (Ed.), *Handbook of noise control*. New York: McGraw-Hill, 1957.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.
- EKMAN, G., FRANKENHAUSER, M., LEVANDER, S., & MELLIS, I. Scales of unpleasantness of electric stimulation. *Scandinavian Journal of Psychology*, 1964, 5, 257-261.
- EKMAN, G., & KÜNNAPAS, T. Measurement of aesthetic value by "direct" and "indirect" methods. *Scandinavian Journal of Psychology*, 1962, 3, 33-39.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- HICKS, C. R. *Fundamental concepts in the design of experiments*. New York: Holt, Rinehart & Winston, 1964.
- KRYTER, K. D. Psychological reactions to aircraft noise. *Science*, 1966, 151, 1346-1355.
- MCGILL, W. J. The slope of the loudness function: A puzzle. In H. Gilliksen & S. Messick (Eds.), *Psychological scaling*. New York: Wiley, 1960.
- SPIETH, W. Annoyance threshold judgments of bands of white noise. *Journal of the Acoustical Society of America*, 1956, 28, 872-877.
- STEVENS, S. S. Cross-modality-validation of subjective scales for loudness, vibration and electric shock. *Journal of Experimental Psychology*, 1959, 57, 201-209.
- STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, 133, 80-86.
- STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *Journal of Experimental Psychology*, 1958, 56, 328-334.
- SULLIVAN, R. Magnitude estimation of radiant heat. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, 3, 115-116. (a)
- SULLIVAN, R. Minimal aversion—thresholds for brightness after light and dark adaptation. *American Journal of Psychology*, 1968, 81, 581-585. (b)
- SULLIVAN, R., WARREN, R., & DABICE, M. Effects of adaptation on judgement of the minimal aversion threshold of white noise. *American Journal of Psychology*, 1969, in press.
- TORGERTSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- WOLFF, B. B. The relationship of experimental pain tolerance to pain threshold: A critique of Gelfand's paper. *Canadian Journal of Psychology*, 1964, 18, 249-253.

(Received August 26, 1967)

PSYCHOLOGY, METAPSYCHOLOGY, AND PSYCHOANALYSIS

PAUL L. WACHTEL¹

New York University

Recent critical examination of psychoanalytic theory has revealed important inadequacies in the energy formulations and in the metapsychology in general. Several authors have suggested that psychoanalytic investigation should be concerned with the purpose or meaning of behavior rather than with the construction of impersonal causal models. It is argued in the present article that *both* purposive and causal perspectives on behavior are essential to a mature psychology. Focus upon the *meaning* of behavior may be an efficient strategy for using the unique data of the psychoanalytic hour, but, for other data, process descriptions often seem more useful. The role of energy constructs is discussed. It is concluded that the lack of measures in psychology quantifiable on an interval scale prevents energy constructs from having the utility they do in physics, but that energy constructs have served to highlight important issues of capacity and limits that must be considered by any meaningful alternative formulation.

Perhaps the most encouraging trend in psychoanalytic theory since the death of Freud has been the tendency to dismantle the iron curtain surrounding its brilliant insights and enabling fruitful intercourse and integrative merger between the observations and viewpoints of psychoanalysts and those of academic psychologists. Such integrative efforts are essential to counter the narrowed perspective and narrowed definition of problems that result from the limited sources of data upon which theories in psychology are generally built. Thus, the structure of psychoanalytic theory, the nature of its assumptions and emphases, derives not only from the nature of Freud's personality and training (Amacher, 1965; Holt, 1965), but also from the fact that its primary data source is the free associations of neurotic patients. Other theories in psychology place similarly heavy emphasis on the bar pressing of rats, the problem solving of college students, etc., and

such emphasis cannot help but constrict the range of questions asked and models proposed.

In the course of the development of psychoanalytic theory, it became the hope of Freud and his followers that psychoanalytic theory would become a general psychology, that it would account not only for neurotic phenomena but for all phenomena in the realm of psychological science. An almost inevitable result of such efforts will be a change in the structure of psychoanalytic theory. As the data base of the theory expands, as it attempts to encompass new observations, to include, for example, the data of laboratory experiments, the theoretical concepts that tie together the observations will require revision. This has been the case generally in the history of science, and even in the history of psychoanalytic theory itself. Thus, as psychoanalytic observation began to focus upon the mechanisms of defense, the topographic model of the conscious, preconscious, and unconscious systems (Freud, 1959) had to be replaced by the structural model of ego, id, and superego; the former model was adequate to account for the early observations but could not survive new data. One can imagine that efforts to expand the explanatory realm of psychoanalytic theory may one day become so successful that the original theory will disappear of its own success. To the extent that the theory

¹ This article was written while the author was a Fellow in the Program for Postdoctoral Study and Research in Psychology and on the staff of the Research Center for Mental Health. The author wishes to express his gratitude to George S. Klein, J. G. Schimek, Linda H. Schoeman, and Philip Bromberg for their valuable comments on earlier drafts of this paper.

Requests for reprints should be sent to Paul L. Wachtel, Research Center for Mental Health, New York University, 4 Washington Place, New York, New York 10003.

truly becomes a general psychology it will cease to be *psychoanalytic* theory, and will tend to be simply psychological theory. The total disappearance of separate theories in psychology and the emergence of a single psychological theory free of the parochial origins of its parts is of course an asymptote from which we are still quite distant, but it is important to keep the emerging goal in mind, for the goal determines what is the essential product of psychoanalytic research.

Freud was far less reverent toward his theoretical constructions than are many of his followers. He recognized (Freud, 1949) quite clearly that it is the *facts* discovered by psychoanalytic work that will stand as his major achievement, rather than the theories. The latter were merely ingenious working principles to tie together the data that had emerged up until then and to point toward new observations. Freud's discoveries regarding unconscious motivation, the expressive meaning of dreams, and the organization of primitive thought will remain valuable long after the words ego, id, or libido have become obsolete. Concern with the distinction between psychoanalytic data and the theory used to account for it has led to increasing critical discussion of the basic theoretical assumptions of psychoanalytic theory in recent years. That aspect of the theory that Freud called his metapsychology has come under the most critical scrutiny. Some of this criticism has been helpful in clarifying assumptions and opening the way for theoretical formulation more suited to our present level of knowledge, but blanket dismissal of all metapsychological perspectives as "reductionistic" (Loevinger, 1966) may obscure some important issues. In particular, it will be suggested below that the energy constructs of Freudian metapsychology, while clearly in need of drastic revision and perhaps even amputation, have served to direct researchers to an important realm of questions. Without careful consideration of the role of energy constructs and other impersonal process descriptions in the theory, there is a danger that a poor formulation may be replaced not by a more adequate one, but by a reluctance to encompass the observations that led to the formulation in the first place.

MEANING AND METAPSYCHOLOGY

Psychoanalytic theory is characterized by two rather separate theoretical approaches. The clinical theory consists of "inferences of *directional* gradients in behavior, and of the *object-relations* involved in these directions [Klein, 1966, p. 10]"; it represents a search for meaning and purpose in behavior. Psychoanalytic metapsychology, on the other hand, reflects Freud's view that purposive concepts are scientifically unacceptable. It is a mechanical quasi-thermodynamic model in the language of forces and energies. Klein (1966) has called it a theory to explain a theory, and has argued that, unlike the clinical propositions, metapsychological concepts are "impervious to the type of data obtainable from the psychoanalyst's hour [p. 15]."

Tension between a humanistic and a mechanistic view of man, between the sense that man is a choosing responsible moral agent and the idea that human actions are as impersonally determined as the motion of the planets, is woven tightly into the fabric of psychoanalysis. One could not expect such tension to be absent from any meaningful psychology, for the paradox of human choice has been a central problem in the study of man for many centuries. Klein (1966) has noted that the impersonal, metapsychological perspective on human behavior has been fallaciously put forth as more *basic* than the purposive explanations of the clinical theory. To understand a man's strivings and feelings, to know him from the inside, is somehow regarded as inferior to observing him as an object in the most literal sense.

Questions of freedom and determinism in psychoanalysis do not directly parallel the conflict between the clinical and metapsychological perspectives. Much of Freud's argument for a determinist position was in the service of building the *clinical* theory, of demonstrating that seemingly meaningless events such as dreams and slips of the tongue could be given meaning if the determining chain of associations were followed back to its source. Nonetheless, the role of metapsychological theorizing is related to the view of human behavior as ultimately determined by an impersonal "natural" process, even if "meanings"

are a convenient description of the intervening events that occur. A brief examination of the psychoanalytic view of determinism and freedom is therefore in order.

FREEDOM AND DETERMINISM

Knight's (1946) discussion of the freedom-determinism issue is probably as close to a standard psychoanalytic position on this issue as one can find. Knight believed that determinism was an essential assumption for a scientific psychology. Yet like many other analysts and therapists (e.g., Mazer, 1960; Wheelis, 1956), Knight recognized that a deterministic viewpoint may be used by patients as resistance to effort at therapeutic change. Many patients claim that they simply cannot help what they are doing, that their behavior is determined and they have no choice or responsibility. Such an attitude may effectively inhibit necessary effort on the part of the patient and may function as a self-fulfilling prophecy that indeed results in the patient remaining the same.

In dealing with this dilemma, Knight emphasized the distinction between determinism as a construct and freedom as a subjective experience. For Knight, freedom is an experience of harmony between one's behavior and one's character, a sense that one is behaving in one's own best interests and is free from compulsion, anxiety, and irrational doubts;

this kind of "freedom" is experienced only by emotionally mature, well-integrated persons, it is the goal sought for one's patients in psychotherapy; and this freedom has nothing whatever to do with free will as a principle governing human behavior, but is a subjective experience *which is itself determined* [italics added, Knight, 1946, p. 372].

Knight claims that such freedom, while subjective, need not be spurious.

The behavior of a well-integrated civilized person can be objectively assessed as "free." Observers see that such a person makes ego-syntonic choices, that his motives are "good," and that he is able to carry out what he wills to do [p. 372].

Nonetheless one is left with the impression from Knight's paper that man's every act and thought is *really* determined, and that freedom, however precious, is finally just an epiphenomenon.

Knight is certainly correct that the term "free will" confuses the issue. One can hardly

use the term without calling up visions of medieval theological debate. But simply to define free will as a subjective experience, causally determined in the same way as the experience of pain upon being burned is determined, seems to be begging the question. Such an approach *assumes* determinism without ever questioning it, and deals with the perplexing issue of freedom by redefining it, and thereby essentially dismissing it.

Free will is a frustrating conception because no one is clear what is meant by "free." It is easy to topple the straw man who speaks of freedom as if it were caprice, of freely willed acts as those that are unrelated to a person's history, character, or environment. But the great hold that the concept of freedom has had on men does not stem from the belief that man's actions are random. It stems from the belief that men's choices are meaningful and that the agony of decision is not simply an interesting phenomenon, to be explained by causal laws external to it, but rather the central fact from which to launch an investigation of man.

In this light, it may be seen that the determinism subscribed to by Knight and other analysts and the freedom proclaimed, for example, by the existentialists are two different perspectives on the same set of observations. Both groups of theorists see men walk and talk and make love and war. In order to organize and make sense of these observations, the determinist views the men he observes as he views all other objects in the universe. They are to him (while he is wearing the cap of the determinist) ultimately understood as bits of moving matter whose motions are described in terms similar in principle (though far more complex and perhaps at a different "level") to those describing the motion of any other particles. From this perspective, free will is indeed simply a "subjective experience," to be explained causally like any other phenomenon.

But there is another perspective on human behavior, another way of organizing the data, which all (healthy?) determinists use in much of their personal lives, and which some psychologists (e.g., the phenomenologists) use in their professional lives as well. Starting not with what is similar between man and the rest

of the universe, but rather with what seems unique to man, one may view men's actions as through their own eyes, one may attempt to construct their phenomenal world, to understand them as men who act and decide on the basis of what they see and feel. From this second perspective (one may call it a "common sense" perspective, though modern psychoanalytic and phenomenological research have carried it far beyond common sense) determinism is neither right nor wrong—it is simply irrelevant, a different way of looking at things. The terms of discourse for such a perspective are feeling and decision rather than cause and effect.

Both perspectives on the facts of human behavior have their utility. As Klein (1966) seems to suggest in a different context, they are mutually incompatible ways of organizing and comprehending observations of human beings (including oneself). Perhaps (though it is doubtful) some day a synthesizing formulation will be created that resolves the contradiction and shows each perspective to be a special case of some general and consistent view. For the present, however, it seems necessary to recognize the need for both perspectives. No man can get along well with just one, and to reduce "psychological science" to those organizations of observations that employ the determinist perspective is simply to impoverish the realm of psychological inquiry. To completely exclude a determinist viewpoint is of course equally productive of sterility.

CAUSAL AND PURPOSIVE PERSPECTIVES

To the extent that metapsychology is, as Klein (1966) suggests, a theory to explain a theory, it reflects a nonrational, quasi-religious commitment to a determinist position. There is false hope (or fear) in the view that we have not *yet* explained all the nuances of human thought, feeling, and behavior in terms of energies and forces. The present author is in agreement with Klein that there are difficulties *in principle* with automatically "converting the terms of clinical observation to impersonal process terms." Neurophysiological or information-processing models, no less than energy models, represent only one perspective on human affairs, and have no

greater claim to describing an *ultimate* truth than do descriptions of striving and meaning. Both perspectives, if used wisely, can provide useful insights.

Psychoanalytic observation is tuned in on a unique level of coherence (Klein, 1966). The structure of the analytic situation facilitates perception by the analyst of patterns of thought organization in the patient, and enables the analyst to notice characteristics of the patient's thinking and behavior that might otherwise go undetected. Certain seemingly unrelated ideas may be expressed in close contiguity with remarkable frequency. Or free association may lead to the expression of ideas and thoughts that are rarely directly expressed at all in any other context. Who would have expected, on the basis of other models of observation, that free-associating patients would so often express incestuous fantasies? On the other hand, psychoanalytic observations also highlights *gaps* in expression not generally as visible elsewhere. The analyst may notice that every time his patient's thoughts seem to be leading toward criticism of an authority, something happens to halt or deflect the progress of his associations. Or he may notice that certain topics, or certain expected reactions, simply never come up as the patient talks about himself and his life day after day.

Such observations, and the generalizations derived from them, are what give psychoanalytic thinking its special cast. Psychoanalysts are trained to notice when a person says he is trying to do one thing and consistently acts as if he were trying to do quite another (or more frequently some odd combination of both). Greater knowledge of the full range of human strivings and of the manifold fears and compromises that bend, disguise, and complicate those strivings is the special fruit of psychoanalytic research. For knowledge of mechanism, on the other hand, and for the building of models based on physical and biological analogies, the data of psychoanalytic sessions are largely irrelevant.

Klein (1966) is careful to point out that impersonal causal models of behavior are not without a place in psychology; they simply are a poor strategy for dealing with the particular opportunities for observation provided

by the psychoanalytic session. Loevinger (1966), however, seems to dismiss metapsychological perspectives more categorically. She is concerned with finding a set of principles that is distinctively psychoanalytic. But what does it mean for a principle to be psychoanalytic? Does it mean that the data upon which it is based came originally from psychoanalytic sessions? Are other sources of data to be dismissed, with psychoanalytic principles remaining "pure" but of little value for answering the more sophisticated questions called for by new observations? Or is a principle psychoanalytic if it is in essence a theoretical statement formulated by Freud or at least an elaboration of such? Then psychoanalytic theory is destined to becoming an interesting historical curio, to be respected for its eminence among early twentieth-century efforts, but of little interest to those concerned with understanding the latest observations of human behavior.

Psychoanalytic insights are far too important to be so embalmed and preserved. The psychoanalytic tradition in psychological thought has been one of the most vital and original. Perhaps the historical accident of its isolation from the academic world and its development into a "movement" contributed to that vitality. But there is now danger that what was once a revolutionary way of thinking is becoming a conservative, consolidated "system," an outcome that is fostered by that same isolation. A psychoanalytic perspective, that is, a view of human behavior based largely on psychoanalytic observations, and emphasizing the role of unconscious motivations, defenses, compromise formations, etc., is still vitally relevant to psychological inquiry. A separate psychoanalytic *theory*, however, at least to the extent that such a theory is not receptive to data from non-psychoanalytic sources, seems increasingly a poor enterprise in which to invest effort. That some metapsychological propositions are "as congenial to many antianalytic psychologies as to psychoanalysis [Loevinger, 1966, p. 433]" is irrelevant. What matters is how congenial they are to the latest data.

Let us consider Loevinger's solution to "one of the most baffling theoretical problems of psychoanalysis [p. 437]," how an impulse

contributes to its own control. She claims that notions such as bound cathexis and counter cathexis add little to the explanation, and suggests instead that "one is controlled, simultaneously or in turn, by one's impulses and one's parents, and one therefore needs to control something in return, if not one's parents, then one's impulses [Loevinger, 1966, p. 437]." This may perhaps be an adequate account of why the developing personality *wants* to control the impulse, but it says nothing about why he *succeeds*. What is lacking in a psychology solely of meanings, as Loevinger calls for, is adequate consideration of *capacities*. As did early psychoanalytic writings, Loevinger's suggested system implies a kind of motivational omnipotence, whereby it is sufficient to explore the *motives* for behavior without considering the *ability* of the individual to efficiently order and execute all the wishes pressing for fulfillment at a given moment. As an example, if a person gets into an accident, it is certainly not always correct to say that he is expressing a wish to hurt himself. A self-destructive wish would no doubt lower the threshold for such events, but it is clear that such wishes are often kept from direct expression by a strong ego, and phenomena such as "distraction" may often play a more important role in accounting for accidents.

Phenomena such as distraction are at the heart of what is missing in Loevinger's scheme. Korchin (1964), for example, has summarized a great deal of evidence suggesting that anxiety may reduce the capacity to "do two things at once." The perceptual-cognitive effects he describes seem vitally relevant in understanding the kinds of behavior that psychoanalysts observe. Certainly, the exact form that behavior in an anxious state takes, as well as the origin of the anxiety itself, is largely a matter of meanings; but without consideration of the constriction of perspective and general narrowing of cognitive organization characteristic of the anxious individual, our understanding of his experience and behavior is incomplete. Similarly, the accounts of individual differences in the organization of attention attempted by Gardner, Klein, and their colleagues (e.g., Gardner, Holzman, Klein, Linton, & Spence,

1959) represent a psychoanalytically oriented approach to character style complementary to and consistent with Freud's discussions of character in terms of libidinal organization. It is a strength rather than a weakness of such work that it is almost impossible to distinguish that part of the conceptual underpinning that comes from psychoanalytic theory and that which comes from other psychological approaches.

COGNITIVE STYLE AND MEANING

The need-in-perception studies were an effort to counter the overly impersonal view of perception prevalent at the time and to highlight the personal meaning of the stimulus for the perceiver. Many of these studies both explicitly and implicitly borrowed from psychoanalytic discoveries of the role of drive and wish in shaping perceptual and cognitive events. As Klein (1958) has pointed out, however, these corrective efforts themselves tended to be one-sided, erring in the direction of overemphasizing the role of drive and disregarding the structural constraints upon drive influence. More precisely, it was the recruiting or meaning-inducing activity of drives that was overemphasized, to the exclusion of the accommodative structures, for Klein's (1958) conception of drive is a unitary one, which emphasizes the coordination of intentions and the means of carrying them out.

One does not understand an individual very well if one knows only the wishes that dominate the recruitment aspects of his personality structure. One must also understand his modes of accommodation and the hierarchic ordering of his wishes and cognitions. Human perception is only sometimes distorted, but it is always selective, and such selection is always limited even if accurate. Regardless of what wishes are currently active, the field-independent individual (Witkin, Dyk, Fatterson, Goodenough, & Karp, 1962) will perceive and experience a different world than the field-dependent, the world of the sharpener will differ from that of the leveler (Gardner et al., 1959), etc. The origin of such differing modes of accommodation may well be largely a matter of the meaning to an individual of certain ways of experiencing the

world.² Such characteristics of the individual's accommodative efforts, however, also enter into his encounters with the world as a kind of automatic filter that he can no longer readily change. Some aspects of personality often designated as styles may better be thought of as capacities than as stylistic preferences (Wachtel, 1968b).

The need for some impersonal process descriptions in personality theory (even for understanding the events of the psychoanalytic hour) is clearer, however, when one considers the somewhat automatic effects of such variables as anxiety upon cognitive functioning. To be sure, the effects of anxiety also are probably subject to individual differences related to the meaning of the anxiety-provoking stimulus. There are indications, for example, that anxiety due to a threat perceived as possible to avert seems to have different consequences than anxiety due to an inevitable harm (Wachtel, 1968a). Nonetheless, the breadth of the individual's cognitive and attentive field does seem to be usefully regarded as a dependent variable in a functional relationship between anxiety and attention (Easterbrook, 1959; Korchin, 1964; Wachtel, 1968a), and breadth of attention is important to consider in understanding phenomena of the therapy hour.

Consider the following clinical example: A female patient, in her early twenties, has a great stake in perceiving herself as a young child, though such an image is, of course, not experienced without considerable conflict. This image of herself helps to reassure her that she is not like her mother, who she perceives as selfish, nasty, and controlling. She lives in a black and white world, where to be a child is to be innocent, pretty, and acceptable, and to be an adult is to be dirty, voracious, and despised.³ Without full awareness, she often behaves in childish ways that call forth responses from others appropriate to

² Shapiro (1965), for example, has vividly illustrated the personal meaning of individual differences in cognitive style.

³ For purposes of exposition, the dynamics of the case must here be greatly simplified. There are, for example, many indications that her identification with her mother is also a cherished key to power, and further that her mother in many respects is also viewed as a kind of ideal.

dealing with someone much younger. Though she complains at times about such treatment, it seems to be quite comfortable for her. On several occasions, she has reported situations in which she was treated with greater respect than usual, and compliments, which she found gratifying, were combined with requests for her to assume responsibilities commensurate with her considerable talents. She found her distress on these occasions puzzling, and it was suggested to her that she would have preferred to receive a simple pat on the head that did not imply that she was an adult capable of making an adult's contribution. On each of these occasions, she became quite agitated and complained that her therapist was talking to her as if she were a child. Similar complaints were made at times about her husband, always at times when he asked her to behave like an adult, for example, by keeping track of her checkbook balance.

Her response to her therapist, as to her husband, obviously had great personal meaning. Threatened with the exposure of and threat to her wish to be a child, she shifted her attention from the content of the therapist's comments to their tone; she further gratified her wish right in the hour by experiencing him as talking to her as if she were a child. But how did she manage to see her therapist in such a fashion? It is not sufficient merely to note that her perception was influenced by the pressure of strong wishes, for as we noted earlier there are considerable constraints upon the distorting effect of wishes. We must assume, I believe, that this patient focused upon some actual features of her therapist's behavior. No doubt her generally childlike manner called forth in her therapist some responses that were more appropriate to addressing a child. The rhythm, tone of voice, or style of phrasing must have provided the cues that led to her perception. That these aspects of the interaction, which meshed so well with her needs, should have become the focus of her attention is not surprising. What is worth noting, however, is the degree to which they remained unmodified by other aspects of the situation. Her field of attention was largely limited to these isolated cues. The larger context in which they occurred, including an explicit message directly

opposite to her perceived meaning (i.e., you were being treated as an adult—people perceive you as an adult), was hardly noticed. In her state of arousal and anxiety, she experienced a narrower range of her therapist's behavior than she normally does, and this narrowing of her field of attention enabled her to focus exclusively on the infantilizing aspect of his remarks. Clearer understanding of the exact nature of the narrowing of attention occurring in anxious states (Wachtel, 1967) would clarify the structural limits within which her motivational and perceptual efforts were expressed, and would suggest what efforts would be needed to correct and amplify her perception of the interpersonal situation. Understanding of the "meaning" of infantilization to her is necessary but not sufficient to guide the therapist's approach to such behavior.

ROLE OF ENERGY CONSTRUCTS

The aspect of psychoanalytic metapsychology that has come under the most intense and persistent criticism has been the economic viewpoint (Rapaport & Gill, 1959), or energy theory. Holt (1967) has pointed out that despite Freud's explicit commitment to physicalistic mechanistic explanation, his concept of psychic energy "is a vitalistic concept in the sense of being similar to and influenced by vital force, and being to a large extent functionally equivalent to it. They are at the least historically and methodologically homologous—buds from the same branch [pp. 24–25]." Holt has discussed in detail the limitations that are thereby inherent in the traditional psychoanalytic energy concept.

It should be noted, however, that the energy concepts of psychoanalysis have had several different functions in the theory. One main function of the energy concept was to deal with the mind-body problem. Freud referred to instincts as a borderland concept between mind and body; and as Holt (1967) points out, Freud often discussed libidinal energy as if it were not simply an abstract psychological concept but rather a form of physical-neural energy, transformed much as heat may be converted to kinetic energy. It is within this framework that loose metaphorical thinking and anthropomorphism have characterized discussion of psychic en-

ergy (libido seen as pushing, pressing for discharge, etc.).

There has been, however, another use of energy concepts by psychoanalytic thinkers that is less subject to the difficulties noted above. The concept of ego energies has been more consistently employed as an abstract nonexistential construct representing the capacity to do (psychological) work, and has thus played a role in theory construction somewhat more akin to the role of energy in physics. As we shall note, this use of an energy concept has its drawbacks too, but examination of the utility of the concept thus far and of the nature of its limitations can pave the way for a better replacement.

Consider the concept of attention cathexis, an aspect of ego energy emphasized by Rapaport (1967). It is based in a general way upon the obvious observation that at any given moment attention to certain aspects of the inner or outer world makes awareness of other aspects less likely. The energy construct is used to coordinate reciprocities among psychological phenomena, relating changes in different parts of the system to each other, much as physical energy is a hypothetical construct simplifying the coordination of reciprocities among physical phenomena. If we divest the concept of the controversial theories of the origins of such energies (cf. Hartmann, Kris, & Loewenstein, 1949; White, 1963), we seem to be left with a useful scheme for relating psychological phenomena, one that may lead to testable hypotheses.

For example, central to the psychoanalytic conception of defense and to the theory of neurosis is the idea that defenses use energy and require a constant effort to prevent the defended-against material from reaching awareness. Rapaport (1951) has described the psychoanalytic conception of consciousness as

a matter of the distribution of attention cathexes, which are available only in a certain quantity. Evidence seems to suggest that these attention-cathexes are identical with those used in counteracting and where excessive energies are required for the latter it limits those available for the former [p. 699].

Gardner et al. (1959) have used this conception in accounting for the empirical relation between the cognitive style of leveling and

repressive defenses, and many neurotic phenomena are also explained psychoanalytically by the impoverishment of the ego, which such diversion of energies implies (Fenichel, 1945). One might expect from this view that if an individual were defending against perceiving an aspect of the stimulus field, his responsiveness to the rest of the field would also be reduced, for part of his capacity to attend is engaged in preventing perception of the offending stimulus.⁴

Some authors, on the other hand (e.g., Kellner, Butters, & Wiener, 1964), have maintained that defenses do not involve a continuing struggle of the sort implied above but rather are a matter merely of the individual's reinforcement history making some responses more likely and others less likely. If this is so, then defending would not consume any of the individual's capacity to attend, and no reduction in overall responsiveness to the field should be evident in those instances of conspicuous inattention to particular features from which defense is inferred. In fact, since some features are avoided, the rest of the field would have less competition for the individual's attention and should be *more* completely perceived. Defense studies that focus not on whether or not a particular brief stimulus is perceived, but rather on the individual's selective attention to a complex field through time, should be able to shed light on the merits of these two competing conceptions, and thereby on the important relation between defense on the one hand and character and neurosis on the other.

It is possible, then, to use energy concepts to guide empirical inquiry. A fair number of studies have originated at least in part from economic concepts (e.g., Lustman, 1957; Schimek & Wachtel, 1968; Schwartz & Schiller, in press; Spence & Greif, 1968; Toman, 1954; Wachtel & Blatt, 1965). The utility of such concepts is considerably limited, however, by the difference in the kinds of quanti-

⁴ The issue is of course far more complicated than can be described in this brief presentation. To maintain and reinforce the defense, for example, some particular aspect of the field might be *more* likely noticed when defending (e.g., loving implications when defending against hostility), but overall awareness of nuances in all aspects of the field might still be reduced.

tative statements made by physicists and psychologists. Energy in physics is measured (indirectly, of course, as in psychology) on an interval scale. There are a precisely specifiable number of footpounds of potential energy embodied in an object suspended above the earth. If that body falls, its velocity upon hitting the ground can be exactly predicted by the formula $1/2 mv^2$ (kinetic energy) = ghM (Potential energy). If these two numbers are not exactly equal, then one knows precisely what portion of the potential energy has been used in moving the resisting air particles, heating the falling object, etc. Since the same "quantity" is presumed to be involved in, for example, mechanical, electrical, and thermal phenomena, a powerful set of equations can be devised to precisely predict changes in seemingly diverse phenomenal realms. There are even "economic" consequences in the more common meaning of the term, for the electric company can bill people for a specific number of kilowatt hours of energy, regardless of whether they heated their house, lit their lamps, cooked a meal, ran electric trains, or brushed their teeth.

The case is quite different in psychology. Here energy is measured on an ordinal scale. One can say only that there is more or less energy used in a given process. To state, for example, the simple functional relationship that the more one exhibits signs of defending against forbidden wishes or ideas, the less one will be attentive to his (internal and external) environment, seems to be exactly equivalent to the proposition that defenses use attentional energy. The formulation in energetic terms adds nothing. This is quite different from the situation in physics, where numerical specification of the quantity of energy involved leads to specific determination of other measures and to the possibility of checking whether all the energy in the system is accounted for. Physicists and engineers would not need an energy concept if the most that could be said was: The more water that falls on the generator, the more houses we can light. It is the possibility of measuring the energy on an interval scale that makes the postulation of such a quantity superior to simple functional description of the relationship between two observable variables. If

one could actually quantify on an interval scale the amount of defensive activity, the amount of perceptual activity, the amount of organizing and structure formation, etc., and show their sum to be a constant, then psychological energy would be an invaluable construct. The day when such measurement is possible, however, seems at this point a matter for psychological science fiction.

At our present level of knowledge, the use of energy constructs may do more harm than good. That the term is used superfluously does not in itself seem a serious matter; it can still be consistently related to observations, and the imagery evoked by the particular metaphor may inspire original hypotheses in some investigators. (Rapaport, 1967, for example, created a complex set of hypotheses that might not have occurred without an energy notion.) It is a serious matter, however, if the illusion of an adequate explanation blinds investigators to the many unanswered questions that remain and delays seeking more precise descriptions of the state of affairs. It is not being unfaithful to an old friend to ask of a theoretical conception: What have you done for me lately? On the contrary, such irreverence is essential for scientific advance. Many concepts are perfectly adequate at an early stage of scientific investigation but may freeze progress if held on to too long. Energy statements have the ring of a finished system, and may perhaps best be held in abeyance until such a system is a more imminent possibility.

If the energy construct is discarded, however, it will be important to bear in mind the reasons for discarding it. In scientific endeavor it is often as important *why* a concept is abandoned as *whether* it is. The difficulties inherent in the use of an energy construct at this stage of theoretical development should not obscure the need for *some* kind of process description and causal analysis in psychological investigation.

The metaphor of a limited quantity of energy that may be deployed in various ways has its problems and limitations, but it does highlight some important issues that it is well to keep in mind. In particular, it is essential to explore the nature of the limits and consequences of human choices, to understand the

price that one pays for committing oneself to a course of action, a style of life, or a mode of perception. The economic point of view emphasizes the limits or boundary conditions of human existence, and the importance of its message should not be dismissed because of the crudity of its formulation. Loewinger (1966) is correct that understanding the meaning of psychological events is an essential aspect of psychoanalytic research, perhaps the most important contribution that such research can make. But unless such understanding is integrated with a firm knowledge of the fabric of constraints within which meaning is given, the psychoanalytic endeavor is not likely to have much meaning at all.

REFERENCES

- AMACHER, M. P. Freud's neurological education and its influence on psychoanalytic theory. *Psychological Issues*, 1964, 4(4), 5-93.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, 66, 183-201.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FREUD, S. An outline of psycho-analysis. (Trans. by J. Strachey) New York: Norton, 1949.
- FREUD, S. The unconscious. In *Collected Papers of Sigmund Freud*. (Trans. by J. Riviere) Vol. 4. New York: Basic Books, 1959.
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. P., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1(4), 1-186.
- HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. Notes on the theory of aggression. *Psychoanalytic Study of the Child*, 1949, 3/4, 9-36.
- HOLT, R. R. A review of some of Freud's biological assumptions and their influence on his theories. In N. S. Greenfield & W. C. Lewis (Eds.), *Psychoanalysis & current biological thought*. Madison: University of Wisconsin Press, 1965.
- HOLT, R. R. Beyond vitalism and mechanism: Freud's concept of psychic energy. In J. Masserman (Ed.), *Science and psychoanalysis*. Vol. 11. New York: Grune & Stratton, 1967.
- KELLNER, H., BUTTERS, N., & WIENER, M. Mechanisms of defense: An alternative response. *Journal of Personality*, 1964, 32, 601-621.
- KLEIN, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of human motives*. New York: Rinehart, 1958.
- KLEIN, G. S. *Two theories or one? Perspectives to change in psychoanalytic theory*. Paper presented at conference of Psychoanalysts of the Southwest, Galveston, Texas, March 1966.
- KNIGHT, R. Determinism, "freedom," and psychotherapy. *Psychiatry*, 1946, 9, 251-262.
- KORCHIN, S. J. Anxiety and cognition. In C. Scheerer (Ed.), *Cognition: Theory, research, promise*. New York: Harper & Row, 1964.
- LOEVINGER, J. Three principles for a psychoanalytic psychology. *Journal of Abnormal Psychology*, 1966, 71, 432-443.
- LUSTMAN, S. Psychic energy and mechanisms of defense. *Psychoanalytic Study of the Child*, 1957, 12, 151-165.
- MAZER, M. The therapeutic function of the belief in will. *Psychiatry*, 1960, 23, 45-52.
- RAPAPORT, D. *The organization and pathology of thought*. New York: Columbia University Press, 1951.
- RAPAPORT, D. The theory of attention cathexis. In M. M. Gill (Ed.), *The collected papers of David Rapaport*. New York: Basic Books, 1967.
- RAPAPORT, D., & GILL, M. M. The points of view and assumptions of metapsychology. *International Journal of Psycho-Analysis*, 1959, 40, 153-162.
- SCHIMEK, J. G., & WACHTEL, P. L. Exploration of effects of distraction, competing tasks and cognitive style on attention deployment. *Perceptual and Motor Skills*, 1969, 28, 567-574.
- SCHWARTZ, F., & SCHILLER, P. A psychoanalytic model of attention and learning. *Psychological Issues*, in press.
- SHAPIRO, D. *Neurotic styles*. New York: Basic Books, 1965.
- SPENCE, D. P., & GREIF, B. Randomness of serial letters as a monitor of attention to distracting prose. Unpublished manuscript, Research Center for Mental Health, New York University, 1968.
- TOMAN, W. Mental or psychic energy and its relation to learning and retention: An experimental contribution. *Acta Psychologica*, 1954, 10, 317-350.
- WACHTEL, P. L. Conceptions of broad and narrow attention. *Psychological Bulletin*, 1967, 68, 417-429.
- WACHTEL, P. L. Anxiety, attention, and coping with threat. *Journal of Abnormal Psychology*, 1968, 73, 137-143. (a)
- WACHTEL, P. L. Style and capacity in analytic functioning. *Journal of Personality*, 1968, 36, 202-212. (b)
- WACHTEL, P. L., & BLATT, S. J. Energy deployment and achievement. *Journal of Consulting Psychology*, 1965, 29, 302-308.
- WHEELIS, A. Will and psychoanalysis. *Journal of the American Psychoanalytic Association*, 1956, 4, 285-303.
- WHITE, R. W. Ego and reality in psychoanalytic theory. *Psychological Issues*, 1963, 3(11).
- WITKIN, H. A., DYK, R., FATERSON, H., GOODENOUGH, D., & KARP, S. *Psychological differentiation*. New York: Wiley, 1962.

(Received July 31, 1968)

MULTIVARIATE CLASSIFICATION OF ALCOHOLICS BY MEANS OF THE MMPI¹

STEVEN G. GOLDSTEIN²

AND

JAMES D. LINDEN

University of Oregon Medical School

Purdue University

The dichotomous conception of different forms of alcoholism generally has led to experimental results where one group was found to be homogeneous and the remainder, as a group, heterogeneous. However, prior research has indicated the existence of many such groups. A multivariate correlation grouping technique was performed on the clinical and validity scales of the MMPI over two large independent samples composed of hospitalized alcoholics. Four alcoholic types were identified and, from actuarial studies of the MMPI, all appeared to be addictive personality patterns. The most common type was characterized by poorly controlled anger and other open forms of emotional expression in frustrating situations.

Most theoretical approaches to the classification of alcoholics are dichotomous (e.g., chronic-acute, process-reactive, etc.) and, as such, suffer from the common failing of any dichotomous explanation: If an occurrence is viewed from one position, events that do not fit the requirements of that position, by the very nature of the dichotomy, must be included in the other (or opposite) position. When applied to a nosological entity this practice generally leads to the identification of at least one heterogeneous group of questionable validity. Jellinek (1952) dismissed the bipolar approach to classification and definition of alcoholics. Instead, he proposed five different types of alcoholism that, for want of a better and more descriptive terminology, he labeled with the first five letters of the Greek alphabet. This classification system admits as applicable all forms of evidence: psychological, physiological, sociological, biological, and nutritional. As such, Jellinek argues, alcoholism is a "true" disease having effects on all areas that comprise the total life space of the individual. This is the only theory of multiple alcoholic types that has persisted even though little research has been accomplished that lends quantitative support to this theory.

¹This research was conducted, in part, through financial support provided by the Scientific Advisory Council to the Licensed Beverage Industries, Inc., while the senior author was at Purdue University.

²Requests for reprints should be sent to Steven G. Goldstein, Department of Medical Psychology, University of Oregon Medical School, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97201.

The majority of previous attempts to classify alcoholics have employed the Minnesota Multiphasic Personality Inventory (MMPI). Reviews of pertinent literature (Sutherland, Schroeder, & Tordella, 1950; Syme, 1957) have evaluated these attempts and have concluded that the lack of any homogeneity in score or profile patterns precluded a quantitative description of the generalized alcoholic personality pattern. Descriptive research with the MMPI using alcoholic Ss seems in agreement on only two concepts: (a) alcoholic MMPI profiles involve high *Pd* scores; and (b) nosologically there seems to be no distinct alcoholic group. In fact, it is probable that there might be several different groups of alcoholics.

Traditionally, research dealing with some theoretical framework is designed with the constraints of the theory in mind. The results of such experiments cannot be clearly interpreted, as experimental elegance becomes victim to the demands of theory that may or may not have been derived originally from empirical tests. The research and resulting theory dealing with addiction do not have a large history of quantitative validation to exemplify the problems that arise from this molar-to-molecular approach. The purpose of this study was to attempt to discover whether quantitative support could be given to a hypothesis of multiple alcoholic types. This was accomplished through the use of a multivariate grouping procedure applied to MMPI profiles obtained from a total sample of 513 male inpatient alcoholics. The major hypothe-

sis tested was that use of the proposed methodology would yield groups within a large subsample of this alcoholic patient sample that could be identified and by MMPI scale scores differentiated from normals. These groups would be reproducible upon replication.

METHOD

MMPI profiles were obtained from 513 behaviorally identified male alcoholics who were tested within 5 days of admission to a state hospital, thus vitiating the problem of the effects of prolonged hospitalization on the test. All Ss had been committed to this state hospital under an act governing treatment of such persons whose alcoholic intake made them unable to function without care. The Ss were assigned randomly to either standardization or replication subsamples prior to scoring. Thirty unanswered questions were used for the cutoff criterion of *S* acceptability. Using this criterion, 239 Ss were assigned to the standardization subsample, and 258 Ss were included in the replication subsample. Sixteen Ss were lost due to an excessive number of unanswered questions or otherwise invalid test results.

The basic problem was to divide *N* Ss into *k* subsets under the restriction that the average of all distances (i.e., profile differences) between Ss within subsets be minimized and the distance between subsets be maximized. Prior research in this area reviewed by Melda and Linden (1965) used distance functions (Nunnally, 1962; Overall, 1964). However, grouping procedures based upon a distance function would appear to have some important drawbacks with regard to the shape of the individual profile and knowledge of the elevation of that profile. While this procedure may be suitable in cases where these variables are relatively unimportant, interpretation of the MMPI necessitates a knowledge of both shape and elevation. For example, an elevated Depression score associated with depressed Hypochondriasis and Hysteria scores in an inverted *V* pattern (i.e., shape) certainly indicates a different type of person than one who produces what has been called the clinical *V* (Dahlstrom & Welsh, 1960).

The procedure used in this study was modification of that originally proposed by Lorr, Bishop, and McNair (1965) and obviates many of the objections raised above. The procedure begins with the construction of a variance-covariance matrix yielding an indication of elevation. From this matrix clusters of Ss are selected whose profile elevation and shape are similar. The first cluster in this procedure is built around a pivot profile that has the highest average correlation with all cluster profiles above a critical limit that is defined as the lower cutting point (*C_L*). *C_L* generally is determined as the .05 level of significance for the Pearson *r* (taking into consideration the *k* variables) less one-half of the associated standard error for that statistic. Retained in this cluster are all profiles that correlate above the *C_L* with the pivot profile and all members initially as-

signed to the cluster. This procedure is preferable to adding profiles that correlate only with the pivot profile, as such profiles may not correlate with other members of the cluster.

An upper cutting point (*C_U*) is determined, for example, as the .05 level of significance for the Pearson *r* plus one-half of the standard error of that statistic. All profiles with an average *r* greater than the *C_U* are removed to prevent clusters from overlapping. This limit also serves to eliminate profiles that are common to more than one cluster. Using the residual *N* a second cluster may be attempted following the same procedure as described for the first cluster. Again, all profiles whose average *r* is above the *C_U* with the second cluster are removed and another clustering procedure initiated for the residual *N*.

A problem occurs here; for example, a cluster made up of two profiles easily might destroy larger clusters because its average correlation with any given cluster might be quite high because of the limited sample size of the cluster. Thus a limit is set on the minimum number of profiles needed to form a cluster. It was suggested by Lorr et al. that the minimum number of profiles required for any cluster be four.

A final check on the procedure may be assessed by recomputing the average correlations between and within clusters and deleting the few profiles that fail to meet the criteria specified above.

This form of correlational clustering was applied independently to both standardization and replication alcoholic substamples for *k* = 15 MMPI scales—the 3 validity scales, 10 clinical scales, Barron's (1956) Ego Strength Scale and Cuadra's (1956) Control in Psychological Adjustment Scale.

RESULTS AND DISCUSSION

Applying the correlational cluster procedure to the standardization sample of 239 hospitalized alcoholics, four types were identified. The member profiles of each type had positive and significant average correlations with other profiles of that type and correlated either negatively or nonsignificantly with members of any of the other three types isolated. One hundred and fourteen Ss (45.42%) of the standardization sample were assigned to one of the four clusters in question. The group of alcoholics (*N* = 251) set aside for replication was clustered using the same constraints as in the case of the standardization group. Again, four types were identified. One hundred and six Ss (42.23%) of the replication sample were clustered by this procedure.

It should be noted that less than one-half of the entire sample was included in the types identified. On the basis of the reported research using this and other grouping proce-

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR CLUSTER STANDARDIZATION
AND REPLICATION GROUPS ACROSS VARIABLES

MMPI scale ^a	Type I		Type II		Type III		Type IV	
	S ^b (N = 46)	R ^c (N = 34)	S (N = 39)	R (N = 44)	S (N = 25)	R (N = 22)	S (N = 4)	R (N = 6)
<i>L</i>								
<i>M</i>	6.83	6.38	3.21	2.86	3.08	2.91	4.00	4.83
<i>SD</i>	2.85	1.81	1.76	1.94	1.47	1.07	1.83	2.04
<i>F</i>								
<i>M</i>	4.59	3.97	11.08	9.27	5.60	5.14	7.25	6.17
<i>SD</i>	2.01	1.73	6.93	3.08	2.35	1.93	3.30	2.32
<i>K</i>								
<i>M</i>	19.96	20.09	10.44	8.89	8.60	11.91	15.50	14.00
<i>SD</i>	3.52	2.60	3.84	3.21	2.99	3.39	4.43	1.79
<i>Hy+5k</i>								
<i>M</i>	13.96	13.56	20.97	16.50	8.80	10.73	14.25	12.67
<i>SD</i>	3.14	3.31	5.41	5.71	2.45	2.62	1.26	3.45
<i>D</i>								
<i>M</i>	21.87	20.59	31.08	27.77	20.40	19.91	17.75	20.00
<i>SD</i>	3.80	4.12	5.25	5.36	3.46	3.39	2.63	4.34
<i>Hy</i>								
<i>M</i>	23.20	23.15	28.23	23.80	14.92	19.45	24.00	16.83
<i>SD</i>	3.70	4.68	6.34	5.85	3.03	3.67	1.83	3.76
<i>Pd+4k</i>								
<i>M</i>	29.74	28.44	29.18	27.93	25.12	24.91	29.50	28.33
<i>SD</i>	4.04	4.08	5.02	4.55	4.12	4.01	2.65	4.80
<i>Mf</i>								
<i>M</i>	20.78	22.65	23.64	24.11	20.36	21.05	23.50	17.33
<i>SD</i>	3.69	3.46	3.83	4.67	3.50	3.39	5.00	2.80
<i>Pa</i>								
<i>M</i>	8.41	8.76	12.72	12.55	8.64	7.18	10.75	8.83
<i>SD</i>	1.88	1.99	4.21	3.05	3.15	2.50	1.26	3.25
<i>Pt+1k</i>								
<i>M</i>	26.52	25.26	37.79	33.59	24.60	23.64	28.25	25.17
<i>SD</i>	4.16	3.01	6.86	5.50	4.51	4.10	4.27	1.94
<i>Sc+1k</i>								
<i>M</i>	24.83	24.18	35.82	30.48	20.84	20.23	27.00	25.17
<i>SD</i>	4.27	3.57	10.24	6.23	5.60	3.45	2.16	2.64
<i>Ma+2k</i>								
<i>M</i>	17.67	17.06	21.85	21.27	20.88	22.23	25.50	23.50
<i>SD</i>	2.63	2.99	5.33	3.48	3.55	3.13	1.29	3.21
<i>Si</i>								
<i>M</i>	22.61	21.29	37.92	35.48	32.28	24.00	24.00	26.83
<i>SD</i>	5.64	5.84	6.55	6.14	5.14	7.17	3.46	4.07
<i>ES</i>								
<i>M</i>	46.00	45.97	32.41	35.86	42.68	47.23	41.50	41.50
<i>SD</i>	3.13	4.37	5.66	5.63	3.33	4.65	4.80	3.89
<i>Cn</i>								
<i>M</i>	21.24	21.21	29.90	29.20	28.00	29.41	24.75	22.50
<i>SD</i>	4.26	3.62	3.66	4.49	3.58	3.51	4.42	4.72

^a Raw score.

^b S = Cluster standardization group.

^c R = Replication group.

dures (e.g., Lorr et al., 1965; Lorr & Radhakrishnan, 1967), this result was anticipated. More of the sample could have been clustered had the limit for association been more liberal. However, lowering the critical values used would have increased the standard error of the mean obtained. Thus, the final decision of what critical value to use was a compromise between the number of Ss

included in four clusters identified and the size of the standard error of the mean. This problem becomes more acute as the number of variables under consideration increases. For this study it was felt that because the similarity of profile shape was important for profile interpretation, the critical level for association should be set at a relatively high level.

TABLE 2

SINGLE CLASSIFICATION ANOVA SYNOPSIS FOR
STANDARDIZATION AND REPLICATION SAMPLES
OF CLUSTER ANALYSIS

MMPI scale	Standardization sample		Replication sample	
	F	p ^a	F	p ^a
L	31.31	.0005	29.24	.0005
F	14.62	.0005	25.11	.0005
K	80.58	.0005	84.42	.0005
Hs	45.91	.0005	8.76	.0005
D	43.73	.0005	19.55	.0005
Hy	36.68	.0005	5.28	.005
Pd	6.10	.001	5.08	.005
Mf	6.06	.001	6.84	.0005
Pa	13.59	.0005	18.50	.0005
Pt	42.75	.0005	36.03	.0005
Sc	27.36	.0005	26.07	.0005
Ma	8.53	.0005	12.02	.0005
Si	55.62	.0005	40.63	.0005
Es	70.77	.0005	36.09	.0005
Cn	48.77	.0005	29.00	.0005

^a $F_{.001} = 4.50$, $df = 3/120$; $F_{.005} = 5.79$, $df = 3/120$; $F_{.0001} = 6.34$, $df = 3/120$.

Table 1 shows the means and standard deviations for the four standardization and replication types. These types are profiled in Figures 1-4. Single classification analyses of variance (ANOVAs) were performed across the four types considering 15 variables for both subsamples. The results of the analyses summarized in Table 2 indicated that there were significant differences between cluster means across all variables.

Newman-Keuls procedures for the identification of significant differences between individual cluster means were performed. The results of these tests indicated that while statistically there were significant differences present, these differences were not consistent for types across variables. As discussed below, the four types identified have been listed separately in volumes of MMPI profiles (e.g., Gilberstadt & Duker, 1965; Hathaway & Meehl, 1951; Marks & Seeman, 1963) and can be described in behavioral terms as being distinct from one another. The profiles included within a type also satisfy the requirement that the average correlation within types across variables deviates from zero-order correlation at or beyond the .05 level of significance. The lack of consistent significant differences among types across variables may be viewed as a difficulty inherent in MMPI profile data. Changes in profile shape and

elevation tend to be confounded through the use of such devices as the Welsh code (Dahlstrom & Welsh, 1960) and thus prevent the production of consistent statistically significant differences when one variable at a time is tested because of shifts in the positions of ordered means and the intersection of profiles on a given variable. Interpretation of the mean profiles as types must be made through the interpretation of the profile as a whole and not through the use of individual variables.

The four alcoholic types identified have distinct dynamic descriptions based on actuarial studies.

Type I

Both Gilberstadt and Duker (1965) and Hathaway and Meehl (1951) list this profile type. Persons presenting such a profile are described as exhibiting a personality trait disturbance (emotionally unstable personality) whose most outstanding trait is a poorly controlled anger that frequently results in "temper tantrums" and other overt forms of emotional expression when presented with frustrating situations. Prior to the revision of the American Psychiatric Association's *Diagnostic Manual*, this profile type was commonly associated with a diagnosis of psychopathic personality, emotional instability. Preu (1944) has pointed out that one of the indications of such a personality disturbance is drug addiction and/or chronic alcoholism.

Type II

An approximation to this profile type is noted in all three actuarial volumes. Generally, the prime nosological term applied to persons presenting such a profile is psychoneurosis. The qualifying description is either anxiety reaction or reactive depression. In either case this appears to be a severe, chronic alcoholism associated with a great number of somatic complaints. Suicidal ideation is evident in the large majority of these cases as well as difficulty in marital relationships characterized by a dominating wife who allows a dependency relationship to become extreme.

Type III

Only Hathaway and Meehl (1951) list this profile. They regard this type, however,

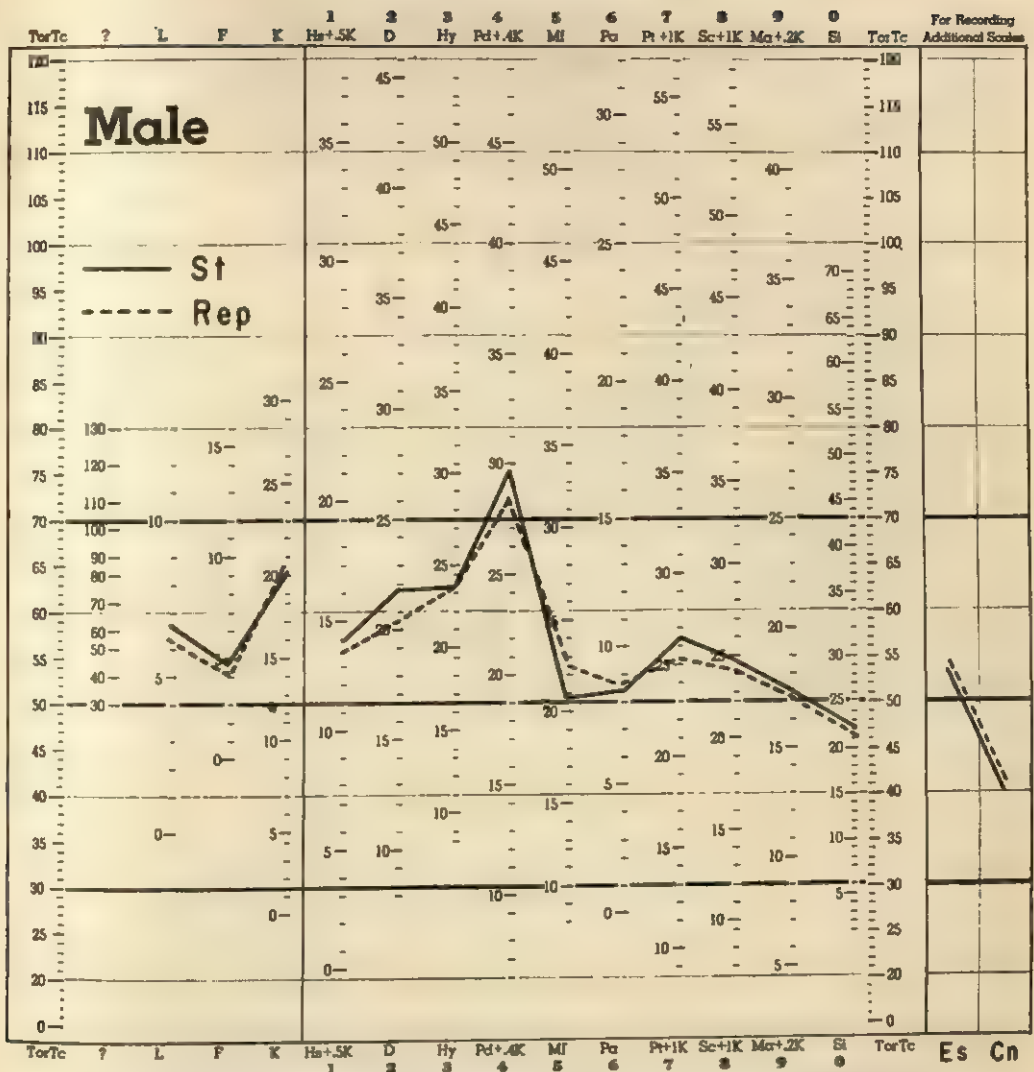


FIG. 1. Standardization and replication profiles for Cluster Type I over MMPI scales.

to be clearly alcoholic and this is their prime diagnosis for such profiles. Secondary to this is the psychopathic personality, mixed-type, classification. Generally, the person presenting such a profile has a long history of chronic alcoholism interspersed with acute alcoholic episodes. He generally exhibits a history of sporadic hospital admissions followed by unsuccessful attempts at abstinence. Length of stay in a hospital is generally short and is marked by the patient's feeling of no need for treatment and an uncooperative patient attitude. Discharge prognosis is usually poor.

Type IV

This code type bears the primary diagnostic classification of alcoholism with secondary characteristics of drug addiction and paranoid features (Hathaway & Meehl, 1951). The person presenting such a profile has used alcohol most of his life in amounts that would be considered excessive in terms of United States normative data. His alcoholic episodes are marked by the addiction of drugs (such as barbiturates) and the substitution of liquids with alcoholic content (such as terpin hydrate or aftershave lotion) for the more

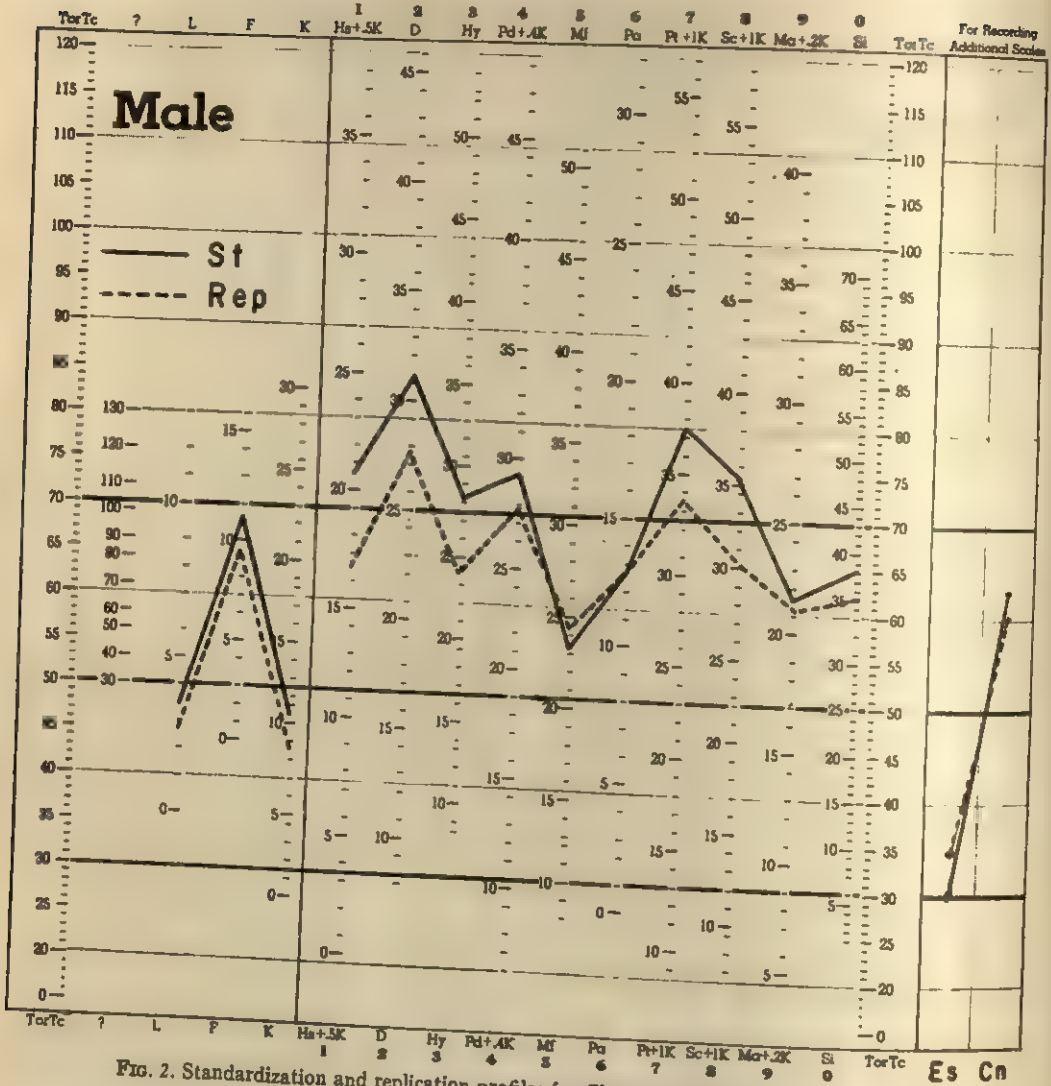


FIG. 2. Standardization and replication profiles for Cluster Type II over MMPI scales.

standard alcoholic beverages. Persons producing a profile such as this have been able to stop drinking completely for indeterminate periods of time and, during these "dry spells," have been known to lead normal and responsible lives.

CONCLUSIONS

The singular concept of an addictive personality has led to the inception of treatment programs, the effects of which have been questionable at best. Because of this, rehabilitative efforts have met with only minimal success. One of the main contentions of this research has been that a group of persons

exhibiting addictive behavior were grossly similar only in overt behavioral symptomology. Further, it was believed that the underlying personality dynamics could vary depending upon the disorder for which the addictive behavior may have been symptomatic. It would appear that this contention has some validity insofar as a group of hospitalized alcoholics was concerned. The dynamics inferred from the MMPI scale scores of persons constituting each of the four alcoholic types identified suggested that such types well might benefit from differential treatment programs designed for the specific type of alcoholic rather than to provide a general

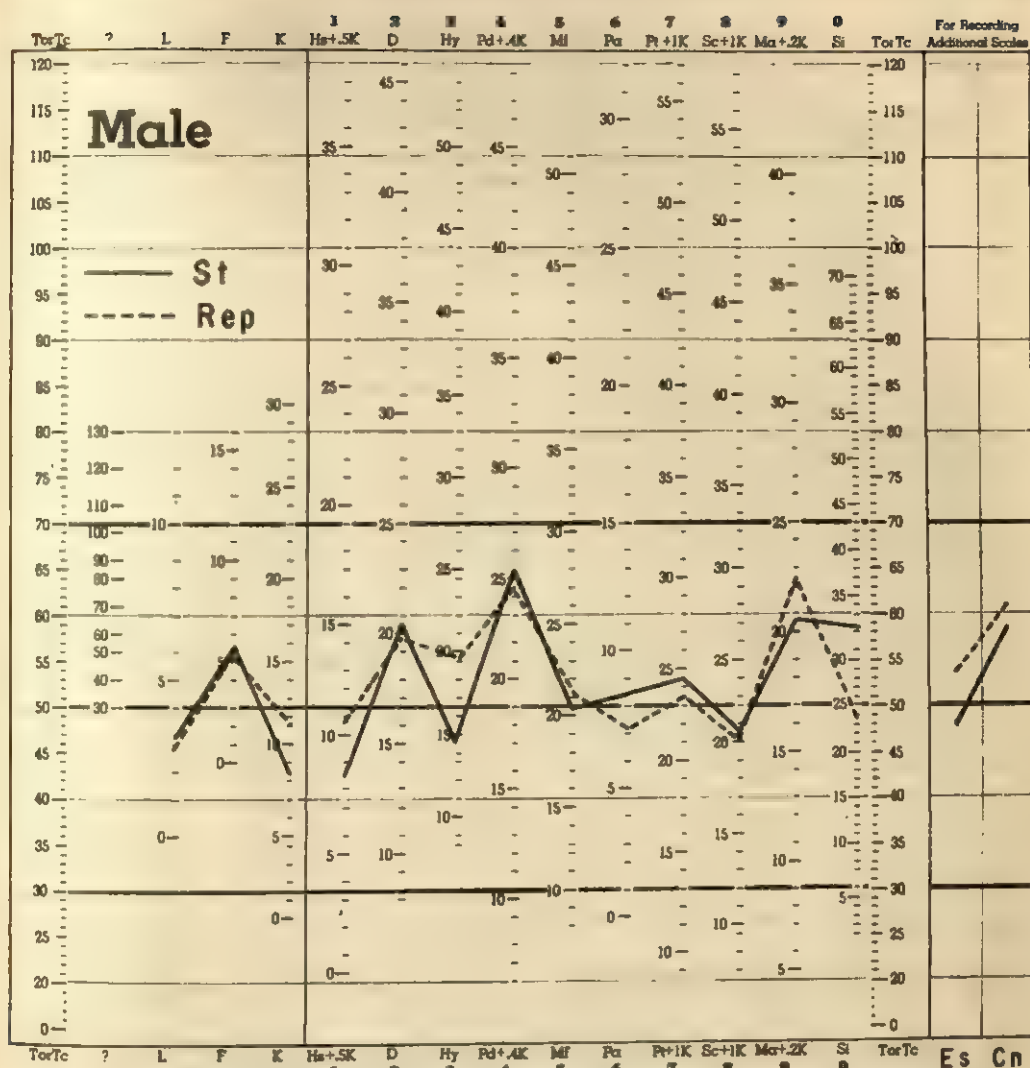


FIG. 3. Standardization and replication profiles for Cluster Type III over MMPI scales.

program for the treatment of all persons called alcoholics.

Independent attempts at further validation by the senior author have suggested that profiles of portions of the group of Ss who were termed Type II degenerate over time into Type I profiles. The reasons for this degeneration are speculative. Both types evidence a relative loss of control in terms of an inability to express emotional content without anger. The Type II profile contains other indications, which generally leads to the diagnosis of some form of psychoneurosis. As these somatic indications respond relatively quickly to treatment, the obtained profile may well

degenerate into a characterological one similar to the Type I profile.

A question may be raised as to the frequency of the existence of the type of alcoholic called here "Type IV." In the total sample of 497 cases, this type of alcoholic profile appeared only 10 times for a relative frequency of occurrence of .02. The fact that it occurred on two successive independent samples producing similar profiles would seem to reject a hypothesis of chance occurrence. On the other hand, mean profiles based upon such small samples generally do not demonstrate reliability. It is the contention of the authors that the successive occurrence of the

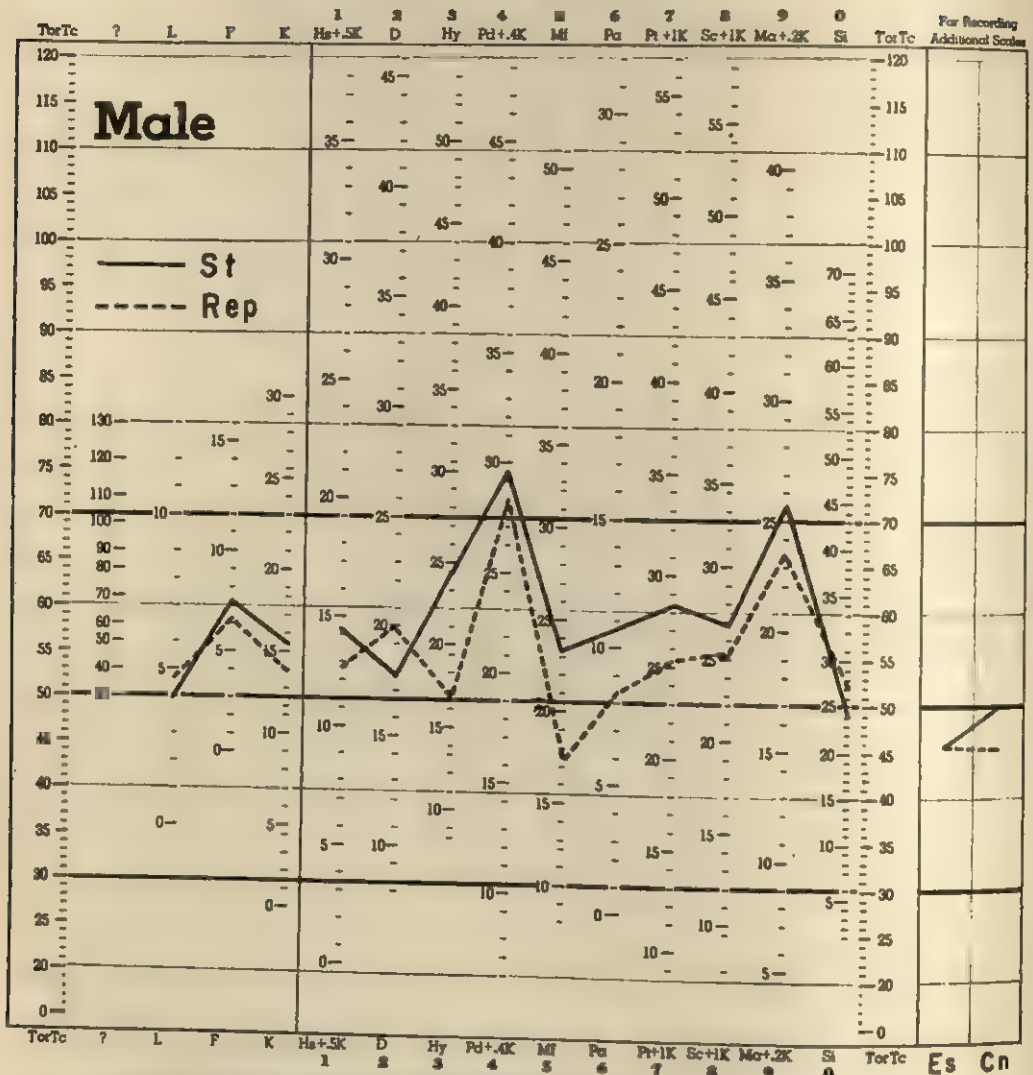


FIG. 4. Standardization and replication profiles for Cluster Type IV over MMPI scales.

Type IV profile is indicative of its existence but this profile type should be viewed and interpreted with caution. It might very well be that this profile is essentially that of the outpatient alcoholic. The type may exist, but with an increase in the size of the sample it may exist in a different profile shape. There is also the possibility that the profile is degenerative; that is, it is actually one of the other three profile types and is indicative of an acute dysfunction that may subside into another type over a relatively short period of time. It should also be noted that only 45% of the standardization and replication

samples could be grouped. Thus the majority of both samples were considered as residual profiles.

The literature contains a number of reported alcoholic patterns that did not appear in the results reported here. Button (1956), Hoyt and Sedlacek (1958), Frame and Osmond (1956), and Rosen (1960), among others, report profiles that show high points on the *D*, *Pd*, and *Ma* scales. While the high points appear in the present results, the remaining portions of the obtained profiles are at variance with these earlier reports. It might be speculated that the discrepancy in

reports is a function of sample selection. However, the present authors felt that this explanation was not adequate to account for all of the discrepancies. It would appear that while univariate approaches used by prior researchers are sufficient to extract high point differences among different subgroups within a sample, they tend not to account for other more subtle profile differences.

This study was directed at the problem of the identification of alcoholics. The samples employed consisted of persons legally committed to a state hospital under the law in effect at that time (1955-57) dealing with the management and treatment of alcoholics. Even though it was felt that the time of testing (within 5 days of entry to the hospital) was optimal in regard to sobriety and lack of institutional effect on the S, the fact that the alcoholic process was deemed serious enough for committal cannot be overlooked. Thus the implications of the results presented here would seem tenuous at best in regard to outpatient alcoholics because of the complex set of interactions between the socioeconomic status of a given person, the severity of the addiction, and the strength of the basic pre-morbid personality structure. Holding the last two constructs constant, the socioeconomic position of a patient alone is enough to determine his disposition in terms of horizontal (outpatient, day, or inpatient care) or vertical treatment attempts (private clinic, sanatorium, or public hospital). With this in mind, it should be realized that the sample of alcoholics used in this study were most likely low on any socioeconomic scale and chronic in terms of recidivist rate. Such persons are not the usual candidates for outpatient care. Therefore, it would seem that the most logical extension of this form of identification and classification would be a replication of the entire study on an outpatient alcoholic sample.

REFERENCES

- BARRON, F. An ego strength scale which predicts response to psychotherapy (Es). In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: University of Minnesota Press, 1956.
- BURTON, A. D. A study of alcoholics with the Minnesota Multiphasic Personality Inventory. *Quarterly Journal of Studies on Alcohol*, 1956, 17, 263-281.
- CUADRA, C. A. A scale for control in psychological adjustment (Cn). In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: University of Minnesota Press, 1956.
- DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: University of Minnesota Press, 1960.
- FRAME, M. C., & OSMOND, W. M. C. Alcoholism, psychopathic personality and psychopathic reactive type. *Medical Proceedings, Johannesburg*, 1956, 2, 257-261.
- GILBERSTADT, H., & DUKER, J. *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders, 1965.
- HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: University of Minnesota Press, 1951.
- HOYT, D. P., & SEDLACEK, G. M. Differentiating alcoholics from normals and abnormals with the MMPI. *Journal of Clinical Psychology*, 1958, 14, 69-74.
- JELLINEK, E. M. Phases of alcohol addiction. *Quarterly Journal of Studies on Alcohol*, 1952, 13, 673-684.
- LORE, M., BISHOP, P. F., & MCNAIR, D. M. Interpersonal types among psychiatric patients. *Journal of Abnormal Psychology*, 1965, 70, 468-472.
- LORE, M., & RADHAKRISHNAN, B. K. A comparison of two methods of cluster analysis. *Educational and Psychological Measurement*, 1967, 27, 47-54.
- MARKS, P. A., & SEEMAN, W. *An atlas for use with the MMPI: Actuarial description of abnormal personality*. Baltimore: Williams & Wilkins, 1963.
- MEIDA, R. J., & LINDEN, J. D. The homogeneous classification of neuropsychiatric patient profiles. In *Proceedings of the 73rd annual convention of the American Psychological Association 1965*. Washington, D. C.: APA, 1965.
- NUNNALLY, J. The analysis of profile data. *Psychological Bulletin*, 1962, 59, 311-319.
- OVERALL, J. E. Note on multivariate methods of profile analysis. *Psychological Bulletin*, 1964, 61, 195-198.
- PREU, P. W. The concept of psychopathic personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 2. New York: Ronald Press, 1944.
- ROSEN, A. A comparative study of alcoholics and psychiatric patients with the MMPI. *Quarterly Journal of Studies on Alcohol*, 1960, 21, 253-266.
- SUTHERLAND, E. D., SCHROEDER, H. G., & TORDELLA, C. L. Personality traits and the alcoholic: A critique of existing studies. *Quarterly Journal of Studies on Alcohol*, 1950, 11, 547-561.
- SYME, L. Personality characteristics and the alcoholic: A critique of current studies. *Quarterly Journal of Studies on Alcohol*, 1957, 18, 288-302.

(Received August 9, 1968)

THE CLINICIAN AS AN ABERRANT ACTUARY: ILLUSORY CORRELATION AND THE INCOMPLETE SENTENCES BLANK¹

B. JAMES STARR² AND EDWARD S. KATKIN

State University of New York at Buffalo

An experiment was conducted in which 60 Ss observed five completed sentences paired equally frequently with each of five statements purporting to describe the major problem of the patient who had completed the paired sentence. The Ss were graduate students in clinical psychology (CGS), graduate students in non-clinical areas of psychology (NCGS), and undergraduate students (US). All groups reported illusory correlations with paralleled relationships generated by a group of experienced clinical psychologists. The groups differed with regard to the degree of confidence which they displayed in their erroneous choices. The US group indicated that they were "sure" and "guessing" more often than the other groups. Reasons for the differential use of confidence categories by the three groups are discussed.

Chapman and Chapman (1967) have pointed to a systematic error of observation that may explain clinicians' continued use of hypotheses and signs which have failed to be validated by research or usage. Chapman and Chapman (1967) described this error, which they have called "illusory correlation," as

the report by an observer of a correlation between two classes of events which in reality (a) are not correlated, or (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction than that which is reported [p. 194].

Collecting data from six experiments they revealed that the clinical signs commonly employed with the Draw-a-Person (DAP) test were discovered anew by naive Ss observing data which contained *no* relationship between drawing characteristics and symptoms. The DAP was chosen for study because of its

wide usage (Sundberg, 1961), despite reported disconfirmation (Swenson, 1957) of many of the clinically relevant signs defined by Machover (1949). The viability of the illusory correlation which they found is attested by the fact that Ss still manifested this observational error even under conditions designed to greatly facilitate the reduction of error.

Although the DAP is a projective technique which has come under severe criticism, the Sentence Completion Test seems to have fared better in the hands of researchers. The Rotter Incomplete Sentences Blank (ISB) has been shown to be a reasonably valid clinical instrument (Rotter & Rafferty, 1950), and it has been employed profitably as a research tool (Goldberg, 1965). The ISB, in common with other sentence completion tests, may derive its validity from the fact that it is a "high-fidelity, narrowband" test (Cronbach, 1960). Furthermore, a survey has revealed that the Sentence Completion Test ranked thirteenth in general usage and second among group tests employed in this country (Sundberg, 1961).

The present study represents an attempt to investigate illusory correlation through the use of completed ISB sentences paired equally frequently with each of a number of statements of patients' problems. The Ss in the Chapman and Chapman (1967) study viewed stimuli consisting of pairs of symptom state-

¹ This research was supported, in part, by National Institute of Dental Research Training Grant 1T1-DE-170 and Research Grant No. MH-11989 from the National Institute of Mental Health, United States Public Health Service. Much of the data analysis and the use of some of the equipment in this investigation were made possible through the aid of the staff and facilities of the Computing Center at the State University of New York at Buffalo. The Computing Center is partially supported by National Institutes of Health Grant FR-00126 and National Science Foundation Grant GP-7318.

² Requests for reprints should be sent to B. James Starr, Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

ments associated with each of a number of drawings. Since the present study employed five descriptive statements paired equally frequently with five completed sentences, the two studies clearly differ with regard to the amount of information that an *S* must process. The Chapmans' *Ss* had to relate qualities which they had abstracted from each of at least 15 drawings to *pairs* of symptom statements. The *Ss* used here had only to observe the same five completed sentences and the same five descriptive statements over and over again in different pairings. Thus, *S* proceeding through the task in the present study had much less new information to process. It was hypothesized that all groups taking part in this experiment (a clinically sophisticated group and two "unsophisticated" control groups) would display the illusory correlation effect. A further hypothesis was that the clinically trained group, with an expectancy for pairs that might "go" together and some confidence in the ISB, would show more confidence in their decisions. Thus, confidence ratings were also obtained in this study.

METHOD

Subjects

Three groups of 20 *Ss* each, all students at the State University of New York at Buffalo, were employed in the present investigation. A group of undergraduate students (US) participated in order to earn credit toward fulfilling one of the requirements of a course in introductory psychology. Groups of graduate students in clinical psychology (CGS) and graduate students in nonclinical areas of psychology (NCGS) also agreed to participate.

Stimuli

Five statements describing the problems of patients were paired with sentences supposedly completed by these patients. Both the descriptive statements and the completed sentences were devised especially for the experiment. The problem descriptions employed were:

1. He has difficulty controlling his aggression.
2. He fears taking positive action.
3. He is disturbed about the strong sexual urges he feels for other men.
4. He feels he has nothing to live for.
5. He complains of perpetual fatigue and illness.

The ISB sentence stems and their completions were:

1. The happiest time—can never be.
2. I regret—nothing I have done.

3. At bedtime—I cry myself to sleep.

4. The best—way to avoid trouble is to do nothing.

5. My greatest fear—is too bad to talk about.

The descriptions and the sentences were typed on IBM punch cards. The descriptive statements were prefixed with the letter "D" followed by a colon, and the completed sentences with "S" followed by a colon. All possible pairs of sentences and descriptions were then combined and stapled together with the descriptions immediately below the completions. Twenty-five different pairs resulted.

Two different random orders of presentation were generated. The final stimulus deck was constituted by placing the pairs in the two random orders together. Each stapled pair was separated from the succeeding pair with a pink IBM card in order to insure the reading of only one pair at a time. The result was 50 pairs of descriptive statements and completed sentences, each pair having an equal frequency of occurrence in the deck (i.e., twice).

Apparatus

Five card stackers were placed on a table in a small room. The stimulus decks were located to the right of the stackers. The instructions and prearranged signals (employed to assure synchronous reading by all *Ss*) were presented on a magnetic tape which had previously been recorded by *E*.

Procedure

On entering the room, a group of *Ss* was asked to seat themselves in front of the stackers. The groups consisted of three, four, or five persons at each testing session. They were told that during the first part of the experiment they would be reading through the decks of cards in front of them. The following instructions were then administered by tape recorder:

As some of you already know, the Rotter Incomplete Sentences Blank is a projective test used to help in psychological assessment. We want to test your powers of observation. You will be shown sentences which have been completed by men with various emotional problems. With each completed sentence you will receive a statement describing what is, in our opinion, the most important problem of the man who has completed this sentence. Many of these men have the same problems. Please study each pair—that is, the completed sentence and the descriptive sentence—very carefully. When you have looked at all of the pairs, I am going to ask you about the sentences completed by men with each type of problem. In front of you there are decks of IBM cards. When you hear the word "Start," remove the rubber band and the first pink card. Study the first pair of sentences. The cards are stapled together. The completed sentences are labeled with an S and the descriptive statements with a D. Do not go on until you hear the word

"Turn." Then remove the next pink card and read the new pair. Read one pair at a time and do not look back at previous pairs. Are there any questions?

The tape recorder was then stopped and relevant portions of the instructions were repeated in answer to any questions which might have arisen. The *E* then said: "Please regulate your reading speed in time with the tape recorder." Following this *E* played the recording of the timed intervals. The *Ss* had 10 sec. in which to read each stimulus pair.

After *Ss* had finished reading the stimulus decks, they were instructed to familiarize themselves with the material on the first page of forms which had been distributed to them. The top line of the form required each *S* to fill in his major field and year level. The form stated:

Some of the sentences were completed by men with the following problem: He has difficulty controlling his aggression. Below are five completed sentences. Choose one which was most often associated with the above problem. Indicate at least a first and second choice. If you feel that some of the sentences were associated equally frequently with this problem, please indicate this. [Immediately below these instructions were the five completed sentences used in the present study. These were followed by further instructions.] Please rate below how sure you are that the above chosen sentence(s) is (are) representative of the data you observed for patients with this type of problem.

positive fairly sure guessing

[After *Ss* had read the form, *E* answered their questions.] There are five pages to this form with two types of items to be answered—a choice of a sentence or sentences and a confidence rating. The confidence rating refers to your first choice or choices. Please be sure to answer both types of items on each page.

Succeeding pages in the form referred to different descriptive statements. The total time for each of the experimental sessions was less than 30 min. A short debriefing session in which *Ss* were told about the experiment and asked not to talk about it followed.

Survey of Clinical Psychologists

In order to facilitate interpretation of any illusory correlation which might emerge, it was necessary to poll a group of clinical psychologists who were familiar with the ISB. The primary purpose of this survey was to ascertain whether *Ss* would "discover" the sentence-description relationships to which experienced clinicians subscribe. The clinicians were given mimeographed forms with the following instructions at the top:

Below are 5 sentences which you should assume have been completed by male patients. The

sentences are followed by five descriptions of the emotional problems of these patients. Please indicate which description best goes with each sentence. Please do not sign your name.

Below the instructions were the five ISB completions and the five descriptive statements employed in this study. At the bottom of the form were two questions designed to assess the utility of the ISB for the respondent. A third question asked the clinician to indicate the number of years in which he had been doing diagnostic work.

Eight doctoral clinical psychologists employed by the State University of New York at Buffalo participated in this phase of the experiment. The anonymity of the respondent was preserved to facilitate the clinician's candor (Chapman & Chapman, 1967). The data obtained from the clinicians remained unanalyzed until after the data collection in other parts of the study was complete.

RESULTS

Response Measures on Clinicians

The clinical psychologists had a mean of 9.25 yr. of experience involving diagnostic work. Completed sentences which the majority of the clinicians saw as being related to each of the descriptive statements are presented in Table 1. The percentage of agreement among members of this group is reported in the second column.

Illusory Correlation in the Experimental Groups

All sentence completions selected as the single "most frequent" correlate for a particular description by at least one-fourth of the *Ss* indicating only one "first" choice are also listed in Table 1. The last column of Table 1 indicates the percentage agreement computed on the data for the combined groups. The first column in that table contains the expected percentages under the null hypothesis. Random distribution of choices (i.e., observed frequencies approximating the expected frequencies) implies an absence of illusory correlation. Cochran *Q* analyses were computed for the descriptive statements to determine if the choices were randomly distributed among the five sentences from which *Ss* had had to pick (Siegel, 1956). All three groups were significantly nonrandom (i.e., consistent) in their choices for the second and fourth descriptions. Beyond this, the third descriptive statement attained signifi-

TABLE 1
PERCENTAGE OF CLINICIANS AND EXPERIMENTAL GROUPS REPORTING THE MOST
POPULAR CORRELATES OF THE FIVE DESCRIPTIVE STATEMENTS

Item	Expected per-centage	Groups				All groups ^b
		Clinicians ^a	CGS ^b	NCGS ^b	US ^b	
Description 1: Difficulty controlling aggression						
Regrets nothing	20	78	29	33	32	31
Avoids troubles	20	11	41*	27	21	29
N		9	17	15	19	51
Description 2: Fears taking action						
Avoids troubles	20	83	61***	67***	67***	65***
N		8	18	15	18	51
Description 3: Strong sexual urges						
Cries to sleep	20	0	41*	27	32	33*
Greatest fear	20	100	41*	33	42*	39*
N		7	17	15	19	51
Description 4: Nothing to live for						
Never happy	20	60	59***	50**	50**	53***
Cries to sleep	20	40	24	38	22	27
N		10	17	16	18	51
Description 5: Perpetual fatigue and illness						
Never happy	20	33	29	25	25	27
Regrets nothing	20	0	35	25	20	27
Cries to sleep	20	50	24	31	25	27
N		6	17	16	20	53

Note.—The criterion for inclusion in this table required that an item be chosen by more than 25% of all the *Ss* included in this analysis. CGS = graduate students in clinical psychology; NCGS = graduate students in nonclinical areas of psychology; US = undergraduate students.

^a Intragroup variation of *N* is due to the clinicians choice of the best combination of sentence and description, irrespective of whether or not an item had previously been used.

^b Intragroup variation of *N* is due to the exclusion of *Ss* who had subscribed to two or more items as equally "most frequent."

* $p < .05$.

** $p < .01$.

*** $p < .001$.

cance for the CGS group and fell just short of significance ($.05 < p < .10$) for the US group. Five *Q* analyses were calculated from the data by combining the three experimental groups. These analyses revealed some consistency in the choice-behavior of the *Ss* for each of the descriptions. The first description attained a $p < .01$; the second, third, and fourth descriptions each attained a $p < .001$; and the fifth description attained a $p < .02$.

Five *Ss*, four in the NCGS group and one in the CGS group, correctly maintained that the five sentences were associated equally frequently with each of the descriptive state-

ments. Three other *Ss*, two in the CGS group and one in the NCGS group, detected the uniform frequencies with regard to one or more (but not all five) of the descriptions. All data in which two or more sentences had been chosen as the most frequent correlates of a description were excluded from consideration in subsequent statistical tests.

Inasmuch as the majority of the *Ss* reported a single sentence as the most common correlate of a particular description, a more specific test of the illusory correlation hypothesis was possible. Since the response measures for these *Ss* were obtained in a forced-choice

TABLE 2

PERCENTAGE OF SUBJECTS REPORTING THE MOST POPULAR CORRELATES OF THE FIVE DESCRIPTIVE STATEMENTS AS A FIRST OR SECOND CHOICE

Item	Ex- pected percent- age	Groups			All groups
		CGS	NCGS	US	
Description 1: Difficulty controlling aggression					
Regrets nothing <i>N</i>	40	47 17	60 15	60 20	56* 52
Description 2: Fears taking action					
Avoids trouble <i>N</i>	40	77** 18	87*** 15	85*** 20	81** 53
Description 3: Strong sexual urges					
Greatest fear <i>N</i>	40	65* 17	60 17	74** 20	67*** 54
Description 4: Nothing to live for					
Never happy Cries to sleep <i>N</i>	40 40	76** 65* 17	76** 76** 17	85*** 55 20	80*** 65*** 54
Description 5: Perpetual fatigue and illness					
Cries to sleep <i>N</i>	40	41 17	69* 16	45 20	51 53

Note.—The sentences included in this table were selected by at least 50% of the *Ss* as a first or second choice. The table also includes those *Ss* who picked two sentences as "most frequent," one of these being a popular correlate. See Table 1 for explanation of abbreviations.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

type of situation any one of the five sentences should be chosen by 20% of the *Ss*. Thus, binomial tests could be applied to these data. The significance levels reported in the middle three columns of Table 1 were based on the binomial test, while those in the last column were based on the Yates corrected z formula for approximating binomial probabilities.

The percentage of *Ss* picking the most popular correlates as either a first or a second choice is shown in Table 2. The objective probability of an *S* selecting any critical sentence as either a first or a second choice is two-fifths as shown in the first column. Given the sample sizes in this study, the z approximation of the binomial test can be applied to

all of the data in Table 2 (Hays, 1963). It may be seen that the data in Table 2 are quite comparable with the data in Table 1. Further, it is apparent that a number of illusory correlations have emerged from the present investigation, and that there is a striking similarity among the clinicians choices and the illusory correlates.

Analysis of Confidence Ratings

Unfortunately, analysis of the confidence ratings was vitiated because most *Ss* avoided the use of extreme categories. However, a chi-square test ($\chi^2 = 14.961$, $df = 4$, $p < .01$) revealed that *Ss* in the three groups did not employ the confidence categories in the same way. The frequency of category choices among the three groups is shown in Table 3. The major component of the chi-square was derived from the use of the "positive" category. The CGS group used this category infrequently—about half as often as the NCGS group. The last-mentioned group, in turn, was much less "positive" than the US group. Another large component of the chi-square can be traced to the relatively infrequent use of the "guessing" category by the US group when compared to the other two groups.

DISCUSSION

The binomial and z tests applied to the data revealed a number of illusory correlations. In addition, collapsing the data from the three experimental groups permitted the inference of a trend toward illusory correlation for each of the descriptive statements used in this study. These inferences are grounded in the fact that *Ss*' choices were significantly nonrandom for each description.

TABLE 3

FREQUENCY OF CATEGORY CHOICES IN CONFIDENCE RATINGS BY EXPERIMENTAL GROUPS

Group	Category		
	Positive	Fairly sure	Guessing
CGS	5	58	37
NCGS	12	52	36
US	21	57	22

Note.—See Table 1 for explanation of abbreviations.

A significant Q in the absence of a specific illusory correlate would seem to indicate that two or more sentences together (neither by itself significant) were most commonly chosen as frequent associates for a particular descriptive statement. Thus, this study confirms the phenomenon described by Chapman and Chapman (1967).

The confirmation of the Chapman and Chapman (1967) findings in the present study underscores the hardy nature of the illusory correlation. Perhaps one of the most striking findings is how few of the 60 Ss said that the sentences appeared with equal frequency. The present investigation also verifies the parallelism between the choices of the experienced clinicians and experimental Ss. Especially supporting here are the responses to the fifth description. Both the clinicians and Ss had difficulty agreeing on a sentence which they could ascribe to this statement.

Perhaps the most parsimonious explanation of the analysis of the confidence ratings would be that the CGS and the NCGS groups have learned to be less extreme in subscribing to "facts." They have probably overlearned the notion that there are few unequivocal psychological relationships. Further, the purposes of the experiment could easily be suspect for members of the CGS group. Having been trained in sentence completion methodology, this group might well have perceived the situ-

ation as evaluative and threatening. If this were so, it might have seemed easier for these Ss to make an erroneous choice to which they were not firmly committed. The Ss might also have felt that an observer would be less critical of S's error, if S had been unsure of himself.

REFERENCES

- CHAPMAN, L. J., & CHAPMAN, J. P. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, 72, 193-204.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper & Row, 1960.
- GOLDBERG, P. A. A review of sentence completion methods in personality assessment. In B. I. Murnstein (Ed.), *Handbook of projective techniques*. New York: Basic Books, 1965.
- HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.
- MACHOVER, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
- ROTTER, J. B., & RAFFERTY, J. E. *Manual: The Rotter Incomplete Sentences Blank College Form*. New York: The Psychological Corporation, 1950.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SUNDBERG, N. D. The practice of psychological testing in clinical services in the United States. *American Psychologist*, 1961, 16, 79-83.
- SWENSON, C., JR. Empirical evaluations of human figure drawings. *Psychological Bulletin*, 1957, 54, 431-466.

(Received August 10, 1968)

VISUAL PERCEPTION OF SEQUENTIAL NUMEROSITY BY NORMALS AND RETARDATES

DONALD H. THOR¹ AND EDWARD A. HOLDEN, JR.²

Edward R. Johnstone Training and Research Center, Bordentown, New Jersey

Fifty-millisecond light stimuli were separated by intervals of 100, 200, 300, 400, 500, and 600 msec. Sequences of 2 to 10 flashes were presented at each inter-stimulus interval for multiple- and single-locus conditions to 20 educable retardates and 20 equal mental age normals. Numerosity discrimination of the younger, normal children was more accurate at all interstimulus intervals ($p < .01$) and at all flash numerosities ($p < .01$). Locus condition interacted significantly with groups ($p < .01$), the normals making more accurate discriminations with single locus presentation over a wider range of interstimulus durations than did the retardates. Locus condition also interacted significantly with numerosity ($p < .05$), normals making more accurate discriminations over a wider range of flash sequences with single locus presentation. The results suggest that retardates process sequential visual stimuli at a slower rate than do normals of equal mental age.

Repetitive visual stimuli, such as flashes of light, must be separated by a minimum duration of time to be perceived as successive; a longer duration is required for accuracy in specifying the number of flashes. Investigations involving intermittent stimulation have been numerous and extensive (Brown, 1965; Misiak, 1967). Variables such as age (Cross, 1963; Misiak & Loranger, 1961), diurnal variation (Walsch & Misiak, 1966), and drug effects (Misiak, Zenhausern, & Salafia, 1966) have been demonstrated to influence critical flicker frequency (CFF).

Retinal location stimulated has been investigated in normal young men by Forsyth and Chapanis (1958) and in brain-injured children by Mark, Meier, and Pasamanick (1958). The latter study was conducted to compare foveal and peripheral (15° nasal) CFF in children with diagnosed pyramidal-tract involvement and normal children. Mean foveal CFF thresholds of the brain-injured group were lower, and magnitude of reduction of threshold values from foveal to peripheral

conditions were significantly greater in the normal group.

The present approach was suggested by the Mark et al. data as well as a more recent study by Spitz and Thor (1968) comparing performance of normals and retardates on a visual backward masking task. The method of multiple sequential stimuli in a numerosity task was imposed to magnify the demand upon the visual processing system and thus possibly reveal discrepant perceptual behavior not observed in a two-flash stimulus presentation task.

Retardates with no evidence of brain damage or other neurological impairment (familial type) and normals of equal mental age were given a sequential numerosity task under two conditions of stimulus presentation: single-locus (SL) and multiple-locus (ML) light flashes. A 2-10 flash range allowed an analysis for difficulty in terms of flash numerosity as well as interstimulus interval (ISI). Although a number of previous studies (Zigler, 1966, 1967) have failed to discriminate between these populations, it was hypothesized that differential performance within groups on the SL and ML conditions would discriminate between the retarded (R) and the equal mental age normals (EMA).

METHOD

Subjects

The Ss were 10 male and 10 female educable retardates (mean chronological age, CA, = 17.66,

¹Requests for reprints should be sent to Donald H. Thor, Johnstone Training and Research Center, Bordentown, New Jersey 08505.

²The authors thank William Gretsche, principal, East Garfield School, Willingboro, New Jersey, and members of his staff for permission and assistance in obtaining normal school children. The authors are also indebted to H. H. Spitz, John J. Winters, Jr., and other members of the Johnstone research staff for their support and critical reading of the manuscript.

$SD = 1.08$; mean $IQ = 68.95$, $SD = 5.62$; mean mental age, $MA = 12.17$, $SD = 1.16$) in residence at the Edward R. Johnstone Training and Research Center and 10 male and 10 female equal MA normals (mean $CA = 10.93$, $SD = .28$; mean $IQ = 116.50$, $SD = 7.05$; mean $MA = 12.73$, $SD = .72$) from a neighboring elementary school.

Inspection of institutional case history files on all retardates revealed no firm neurological evidence of brain damage. In two cases there was suspicion of mild neurological involvement and in one case a history of motor seizures. Typically, the retardates came from impoverished and broken homes, had been neglected in infancy and early childhood, and had long histories of intellectual impairment.

Apparatus and Stimuli

The task required Ss to count the number of light flashes presented successively in sequences varying from 2 to 10. In the SL condition the same light flashed repetitively, whereas in the ML condition sequential flashes were distributed among lamps in different positions. The stimulus display source consisted of a 10×10 matrix of neon lamps ($NE-2s$) spaced 1 in. horizontally and vertically between centers. Horizontal and vertical visual angle between adjacent lamps was $48'$; visual angle of the matrix of lamps was $7^\circ 7'$. The array of lamps was covered with translucent black cloth so that the unilluminated lamps could not be seen. The matrix was mounted vertically with its frontal plane surrounded on the

top and sides by a 3×4 ft. black facade. This composite was set upon a table 31 in. high and was illuminated by a 7½-w. frosted incandescent lamp located 7 ft. in front of its center and 7 ft. above the floor.

Positions of the lamps to be illuminated for each numerosity sequence in the ML condition were selected by use of a random numbers table. In the SL condition the same lamp, approximately in the center of the matrix, was illuminated for all numerosity sequences. Control of numerosity in this condition was effected by manual operation of a switch composite.

Length of ISI was controlled with Gerbands Model 135 recycling timer, and in all presentations each lamp remained illuminated for 50 msec.

A 65-db. white noise background (measured from S 's position relative to the source) effectively masked auditory components of the switching operations.

Procedure

The Ss were tested individually in sessions requiring approximately 40 min. Half of the males and females in each group received the SL condition followed by the ML condition, the remaining Ss receiving these conditions in reverse order. A 10-20 min. rest period intervened between the two halves of the session. Except for the rest period, S sat in a chair 6 ft. in front of the lamp matrix. Prior to each condition S was instructed to count the number of flashes which would appear and report the number observed

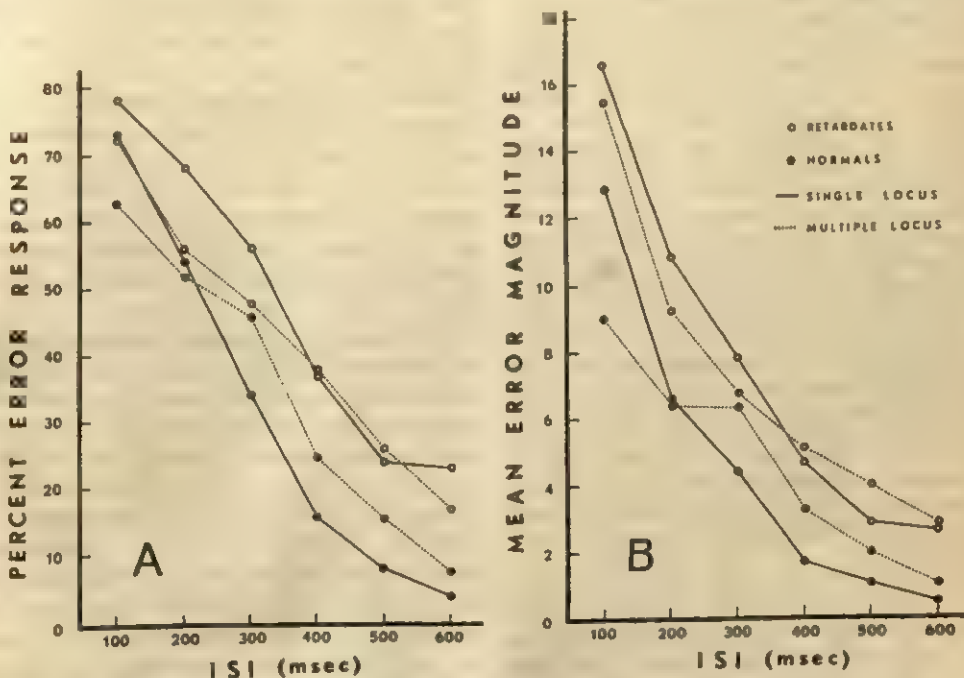


FIG. 1. Error frequency (A) and absolute magnitude of error (B) at varying interstimulus intervals by normals and retardates on single locus and multiple locus viewing conditions.

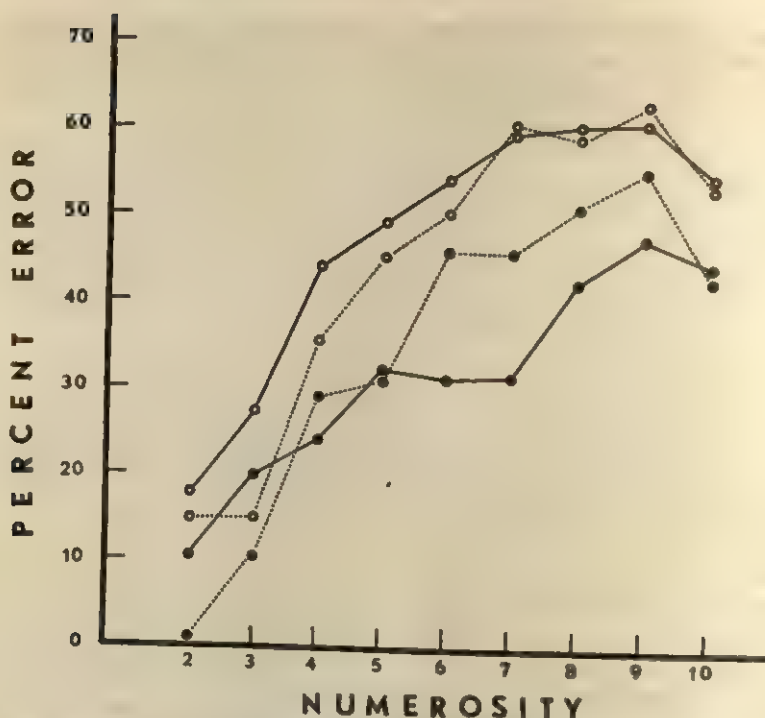


FIG. 2. Error by groups and viewing condition at varying number of light flashes. (Same data points as that of Figure 1.)

when the flashes ceased. Several pretest trials at varying flash numerosities were then given at a 1-sec. ISI to establish that *S* could count and follow instructions. Three retardates were excused from participation for failing this pretest, and another retardate was replaced for having a history of psychosis. After pretesting, *S* was further instructed that every time he gave a correct answer an M&M candy would drop from a dispenser into the receptacle at his side and that all winnings could be retained at the end of the experiment. Retardates were given the alternative of exchanging candy for cigarettes.

Each numerosity sequence was preceded by a 1-sec., 65-db. 500-Hz. warning signal, the termination of which was followed within .6 sec. maximum by the first stimulus light. Under each locus condition all *Ss* received the same order of 54 trials consisting of all possible combinations of numerosities 2, 3, 4, 5, 6, 7, 8, 9, and 10 with ISIs of 100, 200, 300, 400, 500, and 600 msec. Within the 54 trials one randomized order of numerosity was repeated three times, with ISI order varying randomly across the entire trial sequence. Intertrial intervals varied from 3 to 10 sec.

RESULTS

Sex and order variables within the R and EMA groups were nonsignificant and were collapsed for subsequent analyses. In addition, no significant correlations were found between

binocular visual acuity scores and number of correct responses on SL and ML conditions within groups.

Figure 1 describes the decrease in error with increase in ISI by change in percentage of error (A) and by change in mean magnitude of error (B). Percentage of error (sum of errors/total trials $\times 100$) and mean magnitude of error (sum of absolute deviation error/*n*) describe essentially similar relationships.

A three-factor repeated-measures analysis of variance (Winer, 1962, p. 319) was computed with error frequency data and with error magnitude data. Normals made significantly fewer errors ($F = 8.32$, $df = 1/38$, $p < .01$) and errors of lesser magnitude ($F = 7.39$, $df = 1/38$, $p < .01$). Within groups, SL and ML presentations did not differ significantly in error frequency or magnitude. The interaction of Locus Condition \times Groups was significant for error frequency ($F = 7.40$, $df = 1/38$, $p < .01$) but nonsignificant for error magnitude. Expected significance of ISI was apparent with $F = 144.83$ and 159.86 , $df =$

5/190, $p < .001$ for error frequency and magnitude, respectively. ISI also interacted significantly with groups for error magnitude ($F = 3.19$, $df = 5/190$, $p < .01$) but was nonsignificant for error frequency. The Locus Condition \times ISI interaction was significant in both analyses with $F = 6.16$, $df = 5/190$, $p < .001$ for error magnitude and $F = 3.73$, $df = 5/190$, $p < .01$ for error frequency. The triple interaction of Locus Condition \times ISI \times Groups was nonsignificant for the error frequency and significant for error magnitude ($F = 3.02$, $df = 5/190$, $p < .05$), and can be attributed to the greater magnitude of error at the shorter ISIs.

An analysis of variance was also computed to determine the effect of numerosity upon accuracy (Figure 2) and in this case the data were collapsed over ISI and expanded over the nine flash sequences. Groups differed significantly ($F = 8.54$, $df = 1/38$, $p < .01$), retardates making more errors at each numerosity level. The interaction of Groups \times Numerosity was not significant, indicating no group differences in performance with increase in number of flashes. Locus condition was not significant. The interaction of Locus Condition \times Groups was significant ($F = 6.78$, $df = 1/38$, $p < .05$). Numerosity ($F = 71.46$, $df = 8/304$, $p < .001$) and the interaction Locus Condition \times Numerosity ($F = 4.35$, $df = 8/304$, $p < .001$) were both highly significant. The triple interaction was not significant.

Interactions of the single and multiple locus conditions with groups can be seen in the crossing of dashed and solid lines connecting open and closed data points (Figures 1 and 2). More difficult trials at the 100-msec. ISI (Figure 1) reveal a multiple locus contribution to accuracy in both groups that declines with increasing duration of ISI. The point where the SL condition no longer contributes to a greater mean error is reached at approximately 200-msec. ISI by normals and at approximately 400-msec. ISI by the retardates. A similar interaction is also apparent in Figure 2 in terms of numerosity. Normals are more accurate on the SL condition over a greater span of numerosities than are the retardates, who are more accurate on the ML condition.

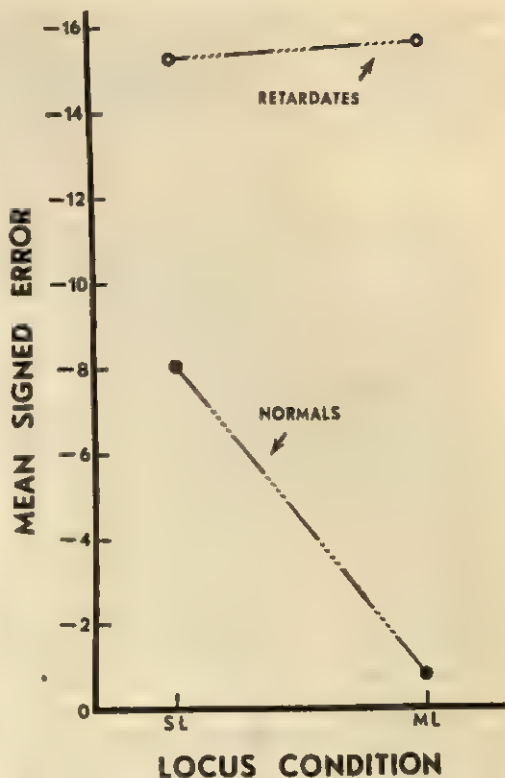


FIG. 3. Direction of error at single and multiple locus viewing conditions by groups. (Algebraic sum of frequency of positive error and frequency of negative error for each S , summated by group and condition, and divided by n .)

Correlations computed with IQ and number of correct responses were $r = .45$, $p < .01$ and $r = .53$, $p < .01$ for the retarded group (Wechsler full scale) on the SL and ML conditions, respectively; for the equal MA group (Kuhlmann-Anderson) $r = .08$, ns (SL), and $r = .44$, $p < .05$ (ML).

Figure 3 describes direction of error for SL and ML condition by R and EMA groups. The difference between positive errors (frequency overestimates) and negative errors (frequency underestimates) were obtained for each S at SL and ML conditions of presentation. A constant was added to eliminate negative numbers and an analysis of variance was computed for Groups \times SL-ML Condition (Winer, 1962, p. 302). Groups differed significantly ($F = 10.88$, $df = 1/38$, $p < .01$) as did the interaction Groups \times Condition ($F = 4.80$, $df = 1/38$, $p < .05$). Normals made sig-

nificantly fewer negative errors, $t = 4.33$, $df = 38$, $p < .001$, than did retardates on the ML condition and the groups did not differ significantly, $t = 1.81$, $df = 38$, $p > .05$, on the SL condition.

DISCUSSION

The gross differential in accuracy of performance between groups was distributed uniformly over all ISI and flash numerosity conditions. On each stimulus condition, the retardates made more errors at every duration and numerosity than did the normals.

Comparison of SL and ML conditions by group mean error and magnitude of error (Figure 1, A and B) demonstrates that normals are able to use SL more accurately than ML over a range of ISI while retardates do not use the SL condition to a distinct advantage at any ISI. The interaction Groups \times SL-ML Condition may be interpreted in reference to a fusion of images when presented in the same locus. If such a fusion occurs at a longer ISI in retardates than in normals, one would expect an interaction similar to that described. That is, the SL condition would contribute to greater error with decreasing ISI sooner for retardates than for normals.

Critical flicker fusion (CFF) has been related to brain damage by Mark, Meier, and Pasamanick (1958). Performance of children with pyramidal-tract damage was compared with that of normals and "the change in threshold from foveal to peripheral conditions discriminated between the two groups ($p \leq 0.01$) [p. 685]." Consistent with the present data was their finding that peripheral threshold decreased in the normals and remained constant in the brain injured.

Relatively greater difficulty with longer sequences of light flashes by retardates does not explain their greater overall error as the interaction Groups \times Numerosity is not significant (Figure 2). It may be noted, however, that the interaction Groups \times Locus Condition over numerosities is significant and comparable with the same interaction over ISI. With longer sequences of light flashes normals make fewer errors in the SL than in the ML condition. Retardates do not reach a point at any numerosity where the SL condition leads to significantly fewer errors.

The significant interaction Groups \times Locus Condition described in Figure 2 is open to several interpretations. The fusion theory is only admissible if numerosities of five or less are discounted since one would expect fusion of 3 and 4 flash sequences on the SL condition as readily as with sequences of 6 or 7 flashes. An alternative may be that normals are more capable of retaining their fixation for a longer duration on a single locus than are the retardates. Another alternative may be that retardates simply do not have a foveal cone of visual acuity (Thomas, 1968, p. 93) comparable with normals. In any case, the significant interaction appears to be due to the atypical curve of the normals on the SL condition.

The change from negative toward positive error in the ML condition by the normals (Figure 3) was further interpreted by inspecting a tabulation of error by distribution according to sign. Normals shifted to positive along the entire distribution, but the total number of trials correct decreased only from 737 to 701. Retardates shifted slightly toward more negative errors on ML, with an increase in total number of trials correct from 563 to 616. Although significant in terms of the 2×2 analysis of variance, change in direction of error showed little effect in the previous analyses for percentage of error and magnitude of absolute error. Accuracy within groups was not influenced by the shift in error direction; thus the normals may be considered as more sensitive to the ML condition in giving higher estimates of flash numerosity in more difficult discriminations than did the retardates. Interpretation of the shift toward higher numerosity estimates for normals under the ML condition is highly speculative, but it can be conjectured that lower counting rates (Vega, 1965) in the retardates handicapped any additional perceptual advantage accruing from the spatially distributed stimuli.

Nonsequential numerosity discrimination in R, equal MA, and equal CA groups has recently been investigated by Spitz, Hoats, and Holden (1968). Dot patterns presented tachistoscopically did not reveal a significant difference in performance between R and equal MA groups although a significant difference was obtained between R and equal CA groups.

Evidently, sequential presentation of stimuli involves behavior that discriminates equal MA and retarded Ss.

In conclusion, retardates appear to differ from equal MA normals in their rate of visual processing of sequential stimuli. The experimental design served to treat the stimulus condition of single or multiple locus as a within-group variable, thus excluding the source of variance attributable to normal-retardate differences. Emotional, motivational, or attentional differences between groups were thus statistically separated from the variables of primary consideration.

Sequential and nonsequential presentation of stimuli to other sense modalities may serve to further delineate behavioral distinctions of retardates and normals. Etiology of such distinctions would appear to be related to central nervous system trauma or malformation in the retarded.

REFERENCES

- BROWN, J. L. Flicker and intermittent stimulation. In C. H. Graham (Ed.), *Vision and visual perception*. New York: Wiley, 1965.
- CROSS, J. P. Relation of age and mental growth to the CFF response in children. *Child Development*, 1963, 34, 739-744.
- FORSYTH, D. M., & CHAPANIS, A. Counting repeated light flashes as a function of their number, their rate of presentation, and retinal location stimulated. *Journal of Experimental Psychology*, 1958, 56, 385-391.
- MARK, H. J., MEIER, P., & PASAMANICK, B. Variability of critical flicker fusion thresholds in brain injured children. *AMA Archives of Neurology and Psychiatry*, 1958, 80, 682-688.
- MISIAR, H. The flicker-fusion test and its applications. *Transactions of the New York Academy of Sciences*, 1967, 29, 616-622.
- MISIAR, H., & LORANGER, A. W. Cerebral dysfunction and intellectual impairment in old age. *Science*, 1961, 134, 1518-1519.
- MISIAR, H., ZENHAUSEN, R., & SALAFIA, W. R. Continuous temporal evaluation of the effect of meprobamate on critical flicker frequency in normal subjects. *Psychopharmacologia (Berlin)*, 1966, 9, 457-461.
- SPITZ, H. H., HOATS, D. L., & HOLDEN, E. A., JR. Numerosity discrimination of tachistoscopically presented dots by mental retardates and normals. *American Journal of Mental Deficiency*, 1968, 73, 127-138.
- SPITZ, H. H., & THOR, D. H. Visual backward masking in retardates and normals. *Perception and Psychophysics*, 1968, 4, 245-246.
- THOMAS, E. L. Movements of the eye. *Scientific American*, 1968, 219, 88-95.
- VEGA, A., JR. Temporal numerosity and the psychological moment: An evaluation. *Dissertation Abstracts*, 1966, 26, 4096.
- WALSH, J. F., & MISIAR, H. Diurnal variation of critical flicker frequency. *Journal of General Psychology*, 1966, 75, 167-175.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- ZIGLER, E. Mental retardation: Current issues and approaches. In L. W. Hoffman & M. L. Hoffman (Eds.), *Review of child development research*. Vol. 2. New York: Russell Sage Foundation, 1966.
- ZIGLER, E. Familial mental retardation: A continuing dilemma. *Science*, 1967, 155, 292-293.

(Received August 26, 1968)

ALCOHOLISM AND CLINICAL SYMPTOMS

ALLEN GOSS¹ AND THOMAS E. MOROSKO

Texas Research Institute of Mental Sciences, Houston

The MMPI was employed in an investigation of the symptom patterns of 200 male alcoholics. The Psychopathic, Depression, and Psychasthenia scales singularly and in combination were the most frequently observed *T* scores with elevations above 70. Indications for drug usage to alleviate clinical symptoms versus techniques to alter drinking behavior were discussed.

Weingold, Lachin, Bell, and Coxe (1968) have reported on the assertion that depression is a characteristic feature of alcoholism. Hypothesizing that alcoholics exhibit clinical signs of depression they selected the Zung (1965) scale as a measure of depression. Also of interest in their study was the evaluation of standard drug therapy effects in reducing depression. Their results indicated that 53 of 73 alcoholic *Ss* exhibited scores which fell above the normal range of Zung's depression scale. The purpose of this study was to investigate symptom patterns of alcoholics with an instrument which permits greater variability and completeness of symptomatic expression.

METHOD

Subjects. The *Ss* were 200 male consecutive admissions to the Texas Research Institute of Mental Sciences Alcoholic Treatment Unit. All of the patients in this sample were outpatients.

Procedure. The *Ss* on the first day at the Institute complete a rather extensive admittance procedure which included psychological testing. The Minnesota Multiphasic Personality Inventory (MMPI), Edwards Personal Preference Schedule (EPPS), Peabody Picture Vocabulary Test (PPVT), Internal-External Scale (I-E), and the Bender-Gestalt Test constituted the psychological test battery. Results from the MMPI receive primary attention in this article.

The population was divided into two groups of 100 each for the data analysis. The sample size was restricted to an *N* of 100 in order to increase the meaningfulness of differences and to have a cross-validation population for comparison purposes.

¹ Also at the University of Saint Thomas and the Baylor College of Medicine. Requests for reprints should be sent to Allen Goss, Texas Research Institute of Mental Sciences, 1200 Moursund Avenue, Texas Medical Center, Houston, Texas 77025.

RESULTS

Table 1 gives age, IQ, and education information for the population. These demographic figures appear to be comparable with other investigations with alcoholic patients (Fitzgerald, Pasewark, & Tanner, 1967; Goss, Morosko, & Sheldon, 1968; Miller, Pokorny, & Hanson, 1968). Table 2 provides mean and *t* ratio MMPI information for the two samples, and Table 3 presents a code analysis of MMPI *T* scores above 70 for single areas as well as for several pattern configurations. The patterns included represent the most frequently observed patterns and pattern combinations. Over half of each population contains patients who exceed a *T* score of 70 on Depression and on Psychopathic. Nearly a third exceed a *T* score of 70 on Hypochondriasis, Hysteria, Psychasthenia, Schizophrenia, and Mania. The most interesting profile pattern is the 427 and the 427 stem combinations. These patterns indicated the presence of a lack of interpersonal warmth, subjective feelings of alienations, depression, and anxiety.

DISCUSSION

The results clearly indicate the presence of multiple symptoms in addition to the speci-

TABLE 1
AGE, IQ, AND EDUCATION INFORMATION
FOR 200 ALCOHOLIC OUTPATIENTS

Group	Age		IQ		Education	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	43.16	8.86	110.67	18.05	10.90	2.87 ^a
2	44.19	10.01	112.49	14.78	11.72	2.72 ^a

Note.—*N* = 100 in each group.

^a These two groups are significantly different at .05 level on education.

TABLE 2

COMPARISON OF MMPI T SCORE FOR TWO GROUPS OF ALCOHOLIC OUTPATIENTS

Scale	Group 1	Group 2	t ratio
Lie	47.73	47.67	.06
Validity	61.24	63.33	1.34
Suppressor	50.14	49.68	.34
Hypochondriasis	62.18	62.90	.32
Depression	72.53	75.22	1.26
Hysteria	62.90	65.24	1.39
Psychopathic	73.09	75.85	1.66
Feminine Interests	59.87	60.05	.12
Paranoia	59.72	62.27	1.55
Psychasthenia	66.84	68.57	.84
Schizophrenia	67.33	65.88	.53
Mania	61.74	62.34	.34
Social Ability	56.58	57.00	.28

Note.—N = 100 in each group.

fied problem of alcoholism. The conclusion of Weingold et al. (1968), which suggests increased use of antidepressant therapeutics with alcoholics, appears questionable in the light of multiple symptoms. If the notion of providing drug therapy for indicated symptoms was followed, antidepressant, antianxiety, antipsychotic, and impulse control compounds would seem necessary for approximately one-fourth (4278) of the population. In the light of recent reviews regarding the effectiveness of drugs with alcoholics, this policy seems questionable (Ditman, 1968; Kissin & Gross, 1968). Pre-post MMPI analysis of alcoholics in a nonmedication treatment program (Sikes, Fabish, & Valles, 1965) indicate that these subjective symptoms decrease without medication.

Since over half of the population exceeds a T score of 70 on the Psychopathic scale a possible conclusion from the MMPI analysis supports an increased use of antabuse (Wallerstein, 1958) with patients who have T scores above 70 on Scale 4. This conclusion makes the assumption that continual drinking is more related to impulse control problems than to feelings of depression. Further work with aversive conditioning procedures also seems promising (Morosko & Baer, 1970). Other possibilities include the use of treatment procedures appropriate for particular subgroups of the alcoholic population. Wallerstein (1958) has reported on the effectiveness of different treatment procedures for particular subgroups of alcoholics. Further

TABLE 3

FREQUENCY OF MMPI T SCORE AREAS AND COMBINATIONS ABOVE 70

T scores and T score combinations above 70	Group 1	Group 2	Total
4	69	58	127
2	64	50	114
24	49	38	87
7	37	35	72
1	34	30	64
3	39	25	64
27	37	27	64
47	33	25	58
8	31	36	57
9	33	24	57
427	32	24	56
49	27	20	47
4278	24	22	46
13	25	20	45
123	26	18	44
6	21	17	38
4271	19	17	36
4273	20	14	34
42781	15	17	32
5	15	16	31
42783	14	14	28
0	14	13	27
19	13	14	27
427813	12	13	25
39	14	10	24
427819	8	9	17
427839]	7	7	14
4278139	7	6	13

work in refining subgroups and increasing the effectiveness of treatment procedures seems essential in the light of the present low success rate with alcoholics.

REFERENCES

- DITMAN, K. S. Review and evaluation of current drug therapies in alcoholism. *Psychosomatic Medicine*, 1966, 28, 667-677.
- FITZGERALD, B. J., PASEWARK, R. A., & TANNER, C. E. Use of the Edwards Personal Preference Schedule with hospitalized alcoholics. *Journal of Clinical Psychology*, 1967, 23, 194-195.
- GOSS, A., MOROSKO, T., & SHELTON, R. Use of the Edwards Personal Preference Schedule with alcoholics in a vocational rehabilitation program. *The Journal of Psychology*, 1968, 68, 287-289.
- KISSIN, B., & GROSS, M. M. Drug therapy in alcoholism. *American Journal of Psychiatry*, 1968, 125, 31-41.
- MILLER, B. A., POKORNY, A. D., & HANSON, P. G. A study of dropouts in an in-patient alcoholism treatment program. *Diseases of the Nervous System*, 1968, 29, 91-99.
- MOROSKO, T., & BAER, P. Case studies in avoidance conditioning of alcoholics. In R. Ulrich, T. Stach-

- nik, & J. Mabry (Eds.), *Control of human behavior*. Vol. 2. Glenview, Ill.: Scott, Foresman, 1970.
- SIKES, M. P., FABISH, G., & VALLES, S. Evaluation of an intensive alcoholic treatment program. *Proceedings of the 73rd Annual Convention of the American Psychological Association*. Washington, D. C.: American Psychological Association, 1965.
- WALLERSTEIN, R. S. Psychologic factors in chronic alcoholism. *Annals of Internal Medicine*, 1958, 48, 114-122.
- WEINGOLD, H. P., LACHIN, J. M., BELL, A. H., & COXE, R. C. Depression as a symptom of alcoholism. *Journal of Abnormal Psychology*, 1968, 73, 195-197.
- ZUNG, W. W. K. A self-rating depression scale. *Archives of General Psychiatry*, 1965, 12, 63-70.

(Received August 26, 1968)

Manuscripts Accepted for Publication in the
Journal of Abnormal Psychology

- Hysterical Personality: An Experimental Investigation of Sex-Role Conflict: Brian T. Jordan* and Bernhard Kempler: St. Tammany Mental Health Center, 403 North Columbia Street, Covington, Louisiana 70435.
- Modification of Placebo Effects by Means of Drugs: Effects of Aspirin and Placebos on Self-Rated Moods: Albert J. Dinnerstein* and Jerome Halm: Bird S. Coler Hospital, Welfare Island, New York, New York 10017.
- Effects of Overtraining on Reversal and Extradimensional Shifts in Schizophrenics: J. Dennis Nolan: Department of Psychology, Ohio State University, Columbus, Ohio 43210.
- Operant Eyelid Conditioning in Trisomy-18: Replication and Extension: Kenneth E. Lloyd,* Harry K. Russell, and Lewis M. Garmize: Department of Psychology, Washington State University, Pullman, Washington 99163.
- Abnormal Social Behavior in Phenylketonuric Monkeys: Arnold S. Chamove, Harry A. Waisman, and Harry F. Harlow*: Primate Research Center, 1223 Capital Court, University of Wisconsin, Madison, Wisconsin 53706.
- Further Validation of a Rorschach Measure of Oral Imagery: A Study of Six Clinical Groups: Lillie Weiss and Joseph Masling*: Department of Psychology, State University of New York, Room 211, Townsend Hall, Buffalo, New York 14214.
- A Pattern Analytic Approach to the Measure of Modes of Expression of Psychological Differentiation: Clyde A. Crego*: 6709 Bracken Court, Springfield, Virginia 22152.
- Psychodynamic Changes through Systematic Desensitization: Leonard J. Kamil*: 1475 Sheridan Avenue, Bronx, New York 10457.
- Case Report: Use of Behavioral Therapy in the Reinstatement of Verbal Behavior in a Mute Psychotic with Chronic Brain Syndrome: Anthony P. Sabatasso and Leonard I. Jacobson*: Department of Psychology, P. O. Box 8185, University of Miami, Coral Gables, Florida 33124.
- Differential Ordering of Stimulus Presentation in Systematic Desensitization: Jon E. Krapfl and M. Mike Nawas*: Department of Psychology, University of Missouri, Columbia, Missouri 65201.
- Treatment Components in Implosive Therapy: Anthony F. Fazio*: Department of Psychology, University of Wisconsin, Milwaukee, Wisconsin 53201.
- Effect of Repetitive Infant-Infant Separation of Young Monkeys: Stephen J. Suomi, Harry F. Harlow,* and Carol J. Domek: Regional Primate Research Center, University of Wisconsin, Madison, Wisconsin 53706.
- Evening Activity, Sleep Mentation, and Subjective Sleep Quality: Peter Hauri*: Department of Psychiatry, Sleep and Dream Laboratory, University of Virginia, Charlottesville, Virginia 22901.

* Asterisk indicates author for whom address is supplied.

DETERMINANTS OF INSTITUTIONAL RELEASE AND PROGNOSIS IN MENTAL RETARDATES¹

WILLIAM VOGEL²

Worcester State Hospital, Worcester, Massachusetts

KAROLY J. KUN AND EDWARD MESHORER

Wrentham Research Foundation, Wrentham, Massachusetts

Familial retardates who were to be released from institutional care (release group, R; $N = 120$) were compared with familial retardates who were not candidates for discharge (institution group, I; $N = 120$) with regard to home backgrounds, family relationships, and preinstitutional and institutional adjustment. Group R presented a preinstitutionalization history of milder retardation than Group I. Group R evidenced cognitive and behavioral improvement from time of admission, whereas Group I evidenced stagnation and deterioration. Group R was subdivided into those who were successful versus unsuccessful in adapting to the community. Prior to institutional release, the successful group evidenced fewer cases of psychiatric disturbance. They evidenced superior performance in community-job situations.

In Windle's (1962) review of the literature with regard to prognosis in mental subnormals, he identified three kinds of variables which are associated with release from institutionalization and/or successful adaptation to the community after release: those which have consistently shown predictive value for successful adjustment after release from institutionalization (e.g., age at release, institutional sanction for release, the attitudes in the home to which the retardate is being released); variables which predict release but which do not seem to be related to successful adjustment after release (e.g., IQ, age at admission, "good" personality, diagnosis); and variables which are commonly assumed to relate to successful adjustment, but which have not been demonstrated to have prognostic value (e.g., length of public institutionalization, organic versus psychosocial etiology, family interest in the retardate, working ability). Research reported since Windle's article generally supports his conclusions (Gage & Wolfson, 1963; Gralicker, Koch, & Henderson, 1965; Strickland & Arrell, 1967), although additional evidence suggests that IQ may have predictive value for adjustment

(Appell, Williams, & Fishell, 1966; Baller, Charles, & Miller, 1967; Jackson & Butler, 1963; Madison, 1964).

Windle noted several methodological problems which make evaluation of the literature difficult: (a) a failure to study homogeneous subgroups of retardates so that meaningful differences are obscured; (b) a failure to employ large enough numbers of Ss so that meaningful generalizations might be made; (c) a failure to employ suitable control samples; (d) inadequate statistical treatment; and (e) concern with individual variables, rather than the predictive value of indexes which employ several variables.

This study is part of a programmatic investigation of factors which are associated with behavioral improvement in institutionalized retardates (Vogel, Kun, & Meshorer, 1967, 1968, 1969). The current report is an attempt to determine, first, in the light of Windle's critique of earlier work, which variables are associated with the administrative decision to release institutionalized retardates into the community. Second, the authors attempt to ascertain which variables are associated with successful adjustment after release.

METHOD

Factors Associated with Institutional Release

Subjects. The first problem was to ascertain the factors associated with institutional release, as con-

¹ This work was supported by Research Grant HD 01798 from the Institute of Child Health and Human Development, United States Public Health Service, to K. J. Kun.

² Requests for reprints should be sent to William Vogel, Department of Psychology, Worcester State Hospital, Worcester, Massachusetts 01604.

trasted with continued institutionalization. Consequently, the last 120 familial retardates were identified who were released with institutional sanction from Wrentham State School into the community (release group, R). Only those Ss were chosen who were released to job placements in the community; none were chosen who were released to parent or foster parent care. These retardates were individually matched, for chronological age (CA) on admission and for current CA, with 120 familial retardates who were not considered to be candidates for release (institutional group, I). The two groups were compared with regard to psychological, social, and biographical variables, described in the following paragraphs. Groups I and R were individually matched for CA and age at admission, first, because the importance of these variables has been amply demonstrated previously (Windle, 1962) and, second, because they are related to other variables (e.g., mental age, MA) to be considered below.

For Group R Ss, intelligence test data were available for the year of admission and for the year of release. For each Group I Ss, similar data were available for the year of admission, and for the year which corresponded to the year of release of the matched partner in Group R.

For the test and retest years, ratings were obtained of the everyday behaviors of Ss, as described in a previous publication (Vogel, Kun, & Meshorer, 1968). In brief, perusal of the records of Wrentham State School suggested that the available behavioral information could be placed into two classifications which were termed personal skills (PS) and social and emotional behavior (SEB). The items categorized as PS reflect acquisition of those perceptual, motoric, and verbal skills which enable a retardate to assume personal responsibility for himself and his possessions and enable him to assist the staff members in caring for the persons and possessions of other retardates. The SEB items reflect acquisition of self-control over emotional expression and ability to relate socially to others as individuals and in groups. Two inventories were constructed, designated PS and SEB, with 15 items in each. Each item was scaled so that a clear order of behaviors was described ranging from absence of a behavioral skill, through partial possession of it, through total mastery. The rater had to decide which of the multiple choices most adequately characterized a retardate's customary level of behavior in respect to that area of activity. One item each from the PS and SEB inventories is shown below; the other 14 items from each inventory were similarly constructed.

Dressing:

- a. Unable to dress self at all.
- b. Can put on simple garments without assistance (e.g., socks, underwear, shorts).
- c. Can dress self, but requires assistance with buttons, zippers, and laces.
- d. Can dress self, including buttons, zippers, and laces without assistance.
- e. Assists in dressing of others.

Relationships to ward personnel:

- a. Actively uncooperative and obstructive.
- b. Passively uncooperative.
- c. Passively cooperative and helpful.
- d. Actively cooperative and helpful.
- e. Helpful in promoting or gaining the cooperation of other retardates with the ward personnel.

The development of the PS and SEB inventories was influenced by the TMR Performance Profile test of DiNola, Kaminsky, and Sternfeld (1965).³ Specifically, their technique was adopted of assessing retardate behavior by means of multiple-choice items arranged to describe a progression of behaviors ranging from no mastery through mastery of a given skill. The published content of the TMR test was generally inapplicable for the authors' purposes. Its detailed scoring requires that the scorer have personal acquaintance with the retardate, rather than the use of institutional records which constituted the source of information in this study. Nevertheless, content from seven TMR items was applicable to the purposes of the present study. The PS and SEB scores discussed and tabulated in the proceeding paragraphs are the result of combining scoring from these TMR items with the items which were prepared.

Two research assistants, trained by the first author, scored the PS and SEB items with an interscorer reliability of .90. Further description of the PS and SEB scales, of their scoring, of the results of a factor analysis of the items, and of interscorer reliability and validity studies are presented in an earlier report (Vogel, Kun, & Meshorer, 1968) in which it was demonstrated that the scoring of the scales was sensitive to, and accurately reflected, changes in the behaviors of subnormals.

Social history. Groups I and R were also compared on a number of variables, reflective of the social, educational, and medical histories of the patients, thought to be relevant in light of Windle's (1962) review. These were (a) socioeconomic status of the parents as reflected in paternal education and occupation, (b) parental interest, as reflected in the number of visits the retardate received and number of visits home, (c) CA at which S's mental deficiency was first noted, (d) the highest educational level achieved by the patient, (e) the highest occupational level at which the patient was able to perform satisfactorily at Wrentham State School, and (f) quality of occupational performance (Ss were rated excellent, satisfactory, or unsatisfactory for the highest occupational level they achieved).

Groups I and R were also compared in regard to psychiatric disability other than mental defect (no psychiatric disturbance, diagnosis of character disorder or neurosis but not psychosis, diagnosis of psychosis). Finally, Groups I and R were compared in regard to sexual composition.

³ The PS and SEB items prepared by the authors and the identification numbers of the relevant TMR items are available from the first author upon request.

TABLE 1

DIFFERENCES BETWEEN RELEASE AND INSTITUTION GROUPS IN INITIAL AND RETEST CA, MA, IQ, PERSONAL SKILLS, AND SOCIAL AND EMOTIONAL BEHAVIORS

Measure	Initial test			Retest		
	Group R	Group I	F	Group R	Group I	F
CA (yr.)	12.3 ± 6.3	12.3 ± 7.4	—	28.5 ± 10	28.5 ± 9.9	—
MA (yr.)	6.55	5.00	27.05*	8.84	5.9	142.1*
IQ	58.7	45.9	60.9*	60.3	41.8	103.0*
SEB	66.8	63.3	ns	84.4	56.1	221.9*
PS	78.1	69.0	25.1*	95.4	79.1	132.58*

Note.—R = release group; I = institution group; PS = personal skills; SEB = social and emotional behaviors.

*p < .001.

Factors Associated with Success after Release

The second problem was to ascertain which factors differentiate retardates who make a successful adjustment to the community from those who must be returned to the institution. Accordingly, from Group R, two subgroups of 30 Ss each were formed, a successful release group (Group Rs) and an unsuccessful release group (Group Ru). Members of Group Rs had successfully maintained themselves outside of the institution for a period of 1 yr., while members of Group Ru had been reinstitutionalized prior to that time. The members of these groups were individually matched for CA, age at admission, year of admission, and year of release. Groups Rs and Ru were compared on initial and retest MA, IQ, PS, and SEB scores, as well as on the social history variables described in the preceding paragraphs.

RESULTS

Factors Associated with Institutional Release

Indexes of adaptive behavior. The results of the *F*-test comparisons between the two groups on initial and retest CA, MA, IQ, PS, and SEB scores are shown in Table 1. In Table 1, it may be seen that MA, IQ, and PS scores differentiate the two groups on initial test at very high levels of statistical significance. However, for each index, the differences between the two groups are much greater on retest. The SEB inventory, which did not differentiate between the groups initially, differentiated on retest to a greater degree than any other index.

In Table 2, it is apparent that on retest there is little overlap between the two groups on the PS and SEB distributions, and little overlap between the two groups on the extremes of the MA and IQ distributions. Thus, in Group R, only 1 S evidenced an MA on retest of less than 6 yr., as contrasted with 51 Ss in Group I; 70 Ss in Group R evi-

denced an MA of 8.1 or greater, as contrasted with 14 Ss in Group I. The overlap between Groups I and R was somewhat greater with IQ than with MA.

On the retest SEB inventory, no Ss in Group R, as contrasted with 67 Ss in Group I, evidenced scores of 60 or less; 79 Ss in

TABLE 2

DISTRIBUTIONS FOR RELEASE AND INSTITUTION GROUPS OF INITIAL AND RETEST MA, IQ, PERSONAL SKILLS, AND SOCIAL AND EMOTIONAL BEHAVIOR SCORES

Measure	Initial test		Retest	
	Group R (N = 120)	Group I (N = 120)	Group R (N = 120)*	Group I (N = 120)*
MA				
0.1-2	1	11	0	7
2.1-4	15	38	0	13
4.1-6	35	29	1	31
6.1-8	34	28	39	46
8.1-10	27	12	42	13
10.1-12	8	2	28	1
IQ				
0-20	1	5	0	9
21-50	23	67	28	80
51-70	81	43	66	27
71-80	14	5	22	3
81-90	1	0	4	1
PS				
0-50	2	19	0	7
51-60	2	9	0	4
61-70	18	32	0	4
71-80	47	29	3	42
81-90	38	23	13	44
91-100	13	8	104	19
SEB				
0-50	5	15	0	45
51-60	26	23	0	22
61-70	46	49	5	24
71-80	39	32	36	23
81-90	4	0	55	6
91-100	0	1	24	0

Note.—See Table 1 for explanation of abbreviations.

* The N for each distribution is 120 Ss, except for MA retest for Groups R and I, where the respective Ns are 110 and 111 Ss.

TABLE 3

PEARSON CORRELATIONS AMONG IQ, MA, CA, PERSONAL SKILLS AND SOCIAL AND EMOTIONAL BEHAVIOR SCORES ON INITIAL TESTING AND ON RETESTING FOR RETARDATES SELECTED FOR RELEASE FROM INSTITUTIONALIZATION

Measure	CA		MA		IQ		SEB	
	Initial	Retest	Initial	Retest	Initial	Retest	Initial	Retest
MA	.73**	-.09						
IQ	-.10	-.27**						
SEB	.13	-.29**	.24**	.28**	.24**	.34**		
			(.22)*	(.26)**	(.25)**	(.28)**		
PS	.65**	-.08	.81**	.41**	.40**	.34**	.44**	.40**
			(.65)**	(.40)**	(.62)**	(.33)**	(.48)**	(.40)**

Note.—See Table 1 for explanation of abbreviations. All correlations are based on $N = 120$, except the retest correlations involving MA, for which $N = 110$. The correlations in parentheses are with CA partialled out.

* $p < .05$.

** $p < .01$.

Group R, as contrasted with only 6 Ss in Group I, evidenced scores of 81 or higher.

On retest on the PS inventory, 104 of 120 Ss in Group R evidenced a score of 91 or better, as contrasted with 19 of 120 Ss in Group I.

The Pearson correlations among IQ, MA, CA, and PS and SEB scores for Group R are shown in Table 3, and those for Group I are shown on Table 4. The "retest" correlations are not based on the raw scores at retesting, but on difference scores from the first to second testing; that is, the retest values in the table represent the intercorrelations of the increments in CA, MA, IQ, PS, and SEB scores from test to retest. The correlations are presented both with and without CA partialled out. The correlations with CA partialled out are shown in parentheses.

Thus presented, Tables 3 and 4 are comparable to previously published tables of inter-

correlations of test and retest MA, IQ, PS, and SEB scores (Vogel, Kun, & Meshorer, 1968). Several observations may be made in this regard. (a) Here, as in the earlier work, PS scores are very highly associated with initial MA scores, even with CA partialled out. (b) Also, as in previous work (Vogel, Kun, & Meshorer, 1968), the correlations of the intelligence measures (MA, IQ) with PS scores are appreciably higher, both on test and retest, than are the correlations with SEB scores. (c) In both groups, the correlations of PS and SEB scores are statistically significant. (d) In both groups the patterns of initial correlations are highly similar in Groups I and R. On retest, the patterns of correlations are again similar, although not to the same extent as on initial test.

Prediction of institutional release from combinations of scores. An attempt was made to determine the extent to which combina-

TABLE 4

PEARSON CORRELATIONS AMONG IQ, MA, CA, PERSONAL SKILLS, AND SOCIAL AND EMOTIONAL BEHAVIOR SCORES ON INITIAL TESTING AND ON RETESTING FOR RETARDATES REMAINING INSTITUTIONALIZED

Measure	CA		MA		IQ		SEB	
	Initial	Retest	Initial	Retest	Initial	Retest	Initial	Retest
MA	.63**	.09						
IQ	.00	-.13						
SEB	.16	-.15	.32**	.42**	.41**	.42**		
			(.29)**	(.44)**	(.42)**	(.41)**		
PS	.57**	-.08	.84**	.78**	.56**	.77**	.55**	.68**
			(.75)**	(.80)**	(.68)**	(.78)**	(.57)**	(.68)**

Note.—See Table 1 for explanation of abbreviations. All correlations are based on $N = 120$, except the retest correlations involving MA, for which $N = 111$. The correlations in parentheses are with CA partialled out.

* $p < .05$.

** $p < .01$.

tions of variables could differentiate, at the time of each retardate's initial admission, those retardates who were eventually to be released from those who were to remain institutionalized.

All Ss in Group R were assigned an arbitrary score of 2, and all Ss in Group I an arbitrary score of 1, so that all 240 retarded Ss were "scored." These scores (1 or 2, for each S) were collectively called group membership scores (GMS). Point biserial correlations were obtained between these GMSs and those variables which significantly ($p < .05$) differentiated the two groups at the time of admission, that is, MA, IQ, SEB, and PS scores, and age at which abnormality was first noted. The correlations were as follows (positive correlations indicate that Group R evidenced higher scores than Group I on a given test): MA \times GMS, .318; IQ \times GMS, .457; SEB \times GMS, .176; PS \times GMS, .317; and Age at Which Abnormality Was First Noted \times GMS, .244.

Using all 240 Ss, a multiple correlation was then obtained between these five variables and GMS. The value of the multiple correlation was .49, $p < .001$. It is apparent that this multiple correlation is not appreciably higher than the correlation between IQ at admission and GMS ($r = .457$).

Social history data. The retardates uniformly were derived from lower-class backgrounds: 94% of 140 fathers of Groups I and R retardates for whom occupational data were available were employed at no higher than the semiskilled level; only two of the fathers had finished high school. There were no differences between Groups I and R in regard to education or occupation.

When the members of the two groups were compared by F test in regard to the CA at which mental subnormality was first noted, the difference was statistically significant ($F = 7.18$, $p < .01$), the mean age for Group R being 4.19 yr., and that for Group I, 3.22 yr.

The retardates were sharply differentiated in regard to accomplishments. When the two groups were combined and divided at the median on the basis of the number of school grades completed, it was found that 73% of Group R but only 32% of Group I completed the second grade ($\chi^2 = 40.1$, $p <$

.001). In regard to occupation, all of the 120 members of Group R, but only 12 of the 120 members of Group I were gainfully employed off the school grounds ($\chi^2 = 192$, $p < .001$).

Finally, the two groups differed significantly in regard to sex; 81 of 120 Ss in Group R but only 60 Ss in Group I were female ($\chi^2 = 6.88$, $p < .01$).

Factors Associated with Success after Release

Neither initial nor retest MA, IQ, PS, or SEB scores differentiated Groups Rs and Ru. The only social history variables which differentiated the groups were psychiatric disability, and job performance.

Of the Ss in Group Rs, 22 had no history of psychiatric diagnosis while 8 had been diagnosed at some earlier period as character disorder, neurotic, or psychotic; the corresponding figures for Ss in Group Ru were 10 and 20 ($\chi^2 = 8.11$, $p < .01$). Thus, Ss who successfully adapted to the community were less likely to evidence a history of psychiatric disorder.

In addition, when Ss' job performances for the highest level jobs they held in the community were rated as excellent, satisfactory, or unsatisfactory all 30 of the Group Rs Ss were rated as excellent, while not one of the Group Ru Ss achieved an excellent rating ($\chi^2 = 56.1$, $p < .001$).

DISCUSSION

Factors Associated with Institutional Release

Differences between Groups I and R on cognitive and adaptive behaviors. Even in a diagnostically homogeneous group of retardates, those Ss chosen for release from the institution (Group R) form a subgroup that is different in character from those who remain institutionalized (Group I). Further, significant group differences are present from the first year of institutionalization, many years prior to release. The initial differences between the two groups became accentuated over time until, on retest, there is little overlap between the groups in terms of their respective distributions of PS and SEB scores, and very little overlap at either extreme of both the MA and the IQ distributions.

The question arises of whether the massive

differences on retest were attributable to Ss being administratively directed on different institutional courses because of staff decisions which were made on the basis of the initial IQ scores. The data do not seem to support this "staff bias" hypothesis. First, at Wrentham State School, all familial retardates, regardless of IQ, are automatically considered at potential "release" candidates, as reflected in the fact that all Ss in this study were assigned to the institutional elementary school. Only 25% of the retardates who are institutionalized at Wrentham State School are so enrolled, representing those retardates who are considered to have the best prognostic potential. Since 32% of the members of Group I completed the second grade or more, they were considered as release candidates for a considerable time.

Second, 20% of the retardates in Group R obtained initial IQ scores of less than 50 so that they would conventionally be considered as "trainable" rather than "educable," that is, incapable of maintaining themselves at large in the community. On the other hand, 40% of Group I retardates initially obtained IQ scores of above 50, so that they would conventionally be considered as the highest quality retardates, ordinarily capable of maintaining themselves in the community. Thus, while Group R evidenced a higher mean initial IQ score than Group I, there were too many individual exceptions to support the hypothesis that the group differences on retest were primarily a reflection of administrative decisions to give favorable treatment to persons with an initial high IQ.

It is also apparent that the differences in institutional adjustment between Groups I and R cannot be attributed to socioeconomic differences in the retardates' home backgrounds. There were no differences between Groups I and R in terms of parental education or occupation. Similarly, the differences cannot be attributed to differences in parental interest after the retardates' institutionalization, as there were no differences between the groups in terms of parental visits received, or in the number of visits to the parental home.

Thus, the differences between the groups on retest probably reflect factors other than

staff bias based on the initial IQ scores, parental interest, or parental socioeconomic status. The data suggest two primary sources of variance.

First, the initial test differences in favor of Group R (high IQ scores, high PS scores, mental retardation not noted until an average of 1 yr. after the mental retardation of the members of Group I became apparent) suggest that a majority of the members of Group R were intellectually and behaviorally less deficient than a majority of the members of Group I from the time prior to institutionalization.

Second, it appears that the members of Group R had a far more favorable adaptation to institutionalization than did the members of Group I. Group I developed far more psychiatric symptomatology, finished fewer school grades, and did much more poorly in terms of occupational accomplishments. As may be seen from Table 1, the members of Group I actually declined in IQ and SEB scores from the time of admission to a period in time which averaged 16 yr. later.

It is not possible, from the current data, to determine the cause of the poor institutional adjustment. It cannot be ascertained, for example, if the members of Group I had a poor educational record and low SEB scores because they developed psychiatric symptomatology; or whether they developed psychiatric symptomatology as a by-product of general neglect and subsequent failure of social and emotional development once they were no longer enrolled in the school program.

The data raise intriguing questions in regard to MA. Only 1 of 120 members of R group evidenced an MA under 6 (see Table 2). Careful inquiry among the administrative and medical staff indicated that this does not reflect an administrative decision; that is, there was no policy of releasing only retardates with an MA of greater than 6. Insofar as intelligence scores had figured in the decision whether or not to release a retardate, it had been the IQ that had been attended to, and even here, no cutoff score had ever been established. As may be seen from Table 2, on retest, 28 of 120 Ss in Group R (that is, 23%) had IQ scores under 50, despite the fact that retardates with IQ scores under 50

are normally considered as incapable of maintaining themselves with any degree of independence outside the institution. Thus, the IQ score had been employed as a general guide for release, rather than as an inviolable criterion. The data raise the question of whether retardates with an MA under 6 are incapable of mastering the skills which serve as a requisite for institutional release, and whether the MA may not be a more valuable index of retardate functioning than IQ. In earlier work it has been suggested that MA relates more successfully to biological and psychological behaviors than does IQ (Vogel & Broverman, 1964) because MA reflects the absolute level of mental ability of an individual, whereas IQ reflects mental ability relative to CA. Consequently, MA should be a better reflection of the absolute level of mental activity and of brain functioning than is IQ.

The most dramatic difference between the groups over the course of hospitalization came not in terms of intellectual functioning, but in terms of SEBs. Although the SEB did not differentiate the two groups on initial test, it was the best single differentiator on retest, there being very little overlap between the two groups. Possibly, the sophisticated skills with which the SEB inventory is concerned are not gained until the later years of institutionalization, so that the device does not successfully differentiate the groups on admission. The PS inventory also sharply differentiated the two groups, on both initial test and retest. These data confirm previous findings (Vogel, Kun, & Meshorer, 1968, 1969) which suggest that PS and SEB scores significantly differentiate groups of retardates along meaningful lines, perhaps to a greater extent than do IQ scores.

It would seem that the SEB scores tap different aspects of retardate development than do the intelligence scores. Thus, on initial test, SEB scores are significantly related to MA in both groups, but the size of the correlations (Group R, .22; Group I, .29) indicates that the two sets of scores have sufficient independence to suggest that they measure different, if related, phenomena.

On the other hand, PS scores are massively correlated with MA on initial test, even with

CA removed (Group R, .65; Group I, .75). In both groups, both initially and on retest, the PS scores have much higher relationships with MA and IQ than do the SEB scores. These interrelationships of PS and SEB scores, on the one hand, to MA and IQ scores, on the other, are highly similar to patterns that have been previously obtained (Vogel, Kun, & Meshorer, 1968). The authors have interpreted them to mean that MA and IQ are related to behaviors which can be gained through instruction (i.e., PS scores) but are less related to behaviors which reflect complex social and interpersonal skills (i.e., SEB scores).

Paradoxically, these data support both those who argue for the efficacy of intelligence tests in evaluating the adaptive potential of retardates (McAndrew & Edgerton, 1964, 1965) and those who argue against (Bauermeister, 1965; Masland, Sarason, & Gladwin, 1958). It is apparent from Table 2 that the MA and IQ scores differentiate the groups at a very high degree of significance on retest, although neither differentiates as well as the SEB inventory.

The finding that, despite individual exceptions, IQ on admission is related to institutional adjustment and to chances of release is in accord with previous research (Windle, 1962). Similarly, the findings that absence of psychiatric illness, good academic records, and vocational placements were associated with institutional release are in accord with previous observations that these are the most common criteria for institutional discharge (Humphreys, 1937; Shafter, 1954; Windle, 1962). The fact that no relationship was observed between parental interest and chance of release is also in support of Windle's (1962) conclusions. Finally, Windle reported that there was no conclusive evidence of a relationship between parental socioeconomic status and chance of release, and suggested that there was a need to investigate this relationship for homogeneous groups of retardates. No relationship between release and socioeconomic status was found.

Factors Associated with Success after Release

It was found that retardates who make a successful adjustment to the community after

their institutional release can be differentiated from those who do not, on the basis of information available prior to institutional release. The retardates who eventually make a successful community adjustment evidence significantly less psychiatric difficulty while still institutionalized at Wrentham. They present a history of superior job performance in the community, as contrasted with the unsuccessful retardates. It seems likely that the two findings are associated, that is, that poor job performance may be a function of a history of psychiatric disorder.

The fact that these two indexes relate to adjustment to the community at such a high level of confidence would appear to be of prime importance. In his review, Windle (1962) reported that the empirical research dealing with work history is scarce, although he cited a study by Shafter (1957) which suggested that custodians' judgments in regard to retardates' work habits were positively related to future adjustment in vocational placement settings.

Overall, the following results of the present study suggest promising areas for future research, especially those which indicate that PS and SEB scores may hold considerable value; the finding that release into the community cannot be expected with an MA of less than 6; the findings that institutional adjustment more than initial intellectual capability or preinstitutional experience may be crucial in determining institutional release; and the findings that Ss who are successful in adapting to the community showed less psychiatric disturbance and superior job performance.

REFERENCES

- APPELL, M. J., WILLIAMS, C. M., & FISHELL, K. N. A residence unit for retarded males: A concluding evaluation. *Training School Bulletin*, 1966, 63, 17-26.
- BALLER, W. R., CHARLES, D. C., & MILLER, E. L. Mid-life attainment of the mentally retarded: A longitudinal study. *Genetic Psychology Monographs*, 1967, 75, 235-329.
- BAUERMEISTER, A. The usefulness of the IQ with severely retarded individuals: A reply to McAndrew and Edgerton. *American Journal of Mental Deficiency*, 1965, 69, 881-882.
- DI NOLA, A. J., KAMINSKY, B. P., & STERNFELD, A. E. *TMR Performance Profile Test*. Ridgefield, N. J.: Reporting Service for Exceptional Children, 1965.
- GAGE, R. M., & WOLFSON, I. N. Four years of experience with a day work program at Litchworth Village. *American Journal of Mental Deficiency*, 1963, 67, 563-568.
- GRALICKER, B. V., KOCH, R., & HENDERSON, R. A. A study of factors influencing placement of retarded children in a state residential institution. *American Journal of Mental Deficiency*, 1965, 69, 553-559.
- HUMPHREYS, E. J. Questionnaires of the 1936-1937 Committee on Research of the American Association on Mental Deficiency. *Proceedings of the American Association on Mental Deficiency*, 1937, 42, 188-222.
- JACKSON, S. K., & BUTLER, A. J. Prediction of successful community placement of institutionalized retardates. *American Journal of Mental Deficiency*, 1963, 68, 211-217.
- MADISON, H. L. Work placement success for the mentally retarded. *American Journal of Mental Deficiency*, 1964, 69, 50-53.
- MASLAND, R. L., SARASON, S. B., & GLADWIN, T. *Mental subnormality: Biological, psychological and cultural factors*. New York: Basic Books, 1958.
- MCANDREW, C., & EDGERTON, R. IQ and the social competence of the profoundly retarded. *American Journal of Mental Deficiency*, 1964, 69, 385-389.
- MCANDREW, C., & EDGERTON, R. Reply to Bauermeister. *American Journal of Mental Deficiency*, 1965, 69, 883.
- SHAFTER, A. J. The vocational placement of institutionalized mental defectives in the United States. *American Journal of Mental Deficiency*, 1954, 59, 279-307.
- SHAFTER, A. J. Criteria for selecting institutionalized mental defectives for vocational placement. *American Journal of Mental Deficiency*, 1957, 61, 599-616.
- STRICKLAND, C. G., & ARRELL, V. M. Employment of the mentally retarded. *Exceptional Children*, 1967, 34, 21-24.
- VOGEL, W., & BROVERMAN, D. M. Relationship between EEG and test intelligence: A critical review. *Psychological Bulletin*, 1964, 62, 132-144.
- VOGEL, W., KUN, K. J., & MESHORER, E. The effects of environmental enrichment and environmental deprivation on cognitive functioning in institutionalized retardates. *Journal of Consulting Psychology*, 1967, 31, 570-576.
- VOGEL, W., KUN, K. J., & MESHORER, E. Changes in adaptive behavior in institutionalized retardates in response to environmental enrichment or deprivation. *Journal of Consulting and Clinical Psychology*, 1968, 32, 76-82.
- VOGEL, W., KUN, K. J., & MESHORER, E. The behavioral significance of EEG abnormality in mental defectives. *American Journal of Mental Deficiency*, 1969, 74, 62-68.
- WINDLE, C. Prognosis of mental subnormals. *American Journal of Mental Deficiency*, 1962, 66(5, Pt. 2).

(Received September 11, 1968)

EFFECT OF HYPNOTIC AGE REGRESSION ON THE MAGNITUDE OF THE PONZO AND POGGENDORFF ILLUSIONS¹

MICHAEL PARRISH, RICHARD M. LUNDY,² AND HERSCHEL W. LEIBOWITZ

Pennsylvania State University

The magnitudes of the Ponzo and Poggendorff illusions among 10 hypnotically age-regressed college students were typical of normative data obtained from young children. This effect was almost complete for the Ponzo but not as pronounced for the Poggendorff illusion. A control group under task motivation instructions failed to produce comparable results. It appears as if age regression facilitates the use or nonuse of visual cues and mechanisms in the manner typical of earlier stages of perceptual development. It is suggested that these illusions may serve as an objective indicator of hypnotic age regression, and that hypnotic age regression may provide a methodological approach to the experimental study of perceptual development.

The use of standard-hypnotic-induction techniques has been reported to facilitate the retrieval of responses, verbal and nonverbal, presumably not emitted for long periods of time. The retrieval of past events by means of hypnosis is usually accomplished by suggesting that *S* is actually the age at which the event is assumed to have occurred. This technique, referred to as "age regression," has been applied to the retrieval of memory for specific events, such as dates (True, 1949), performance on intelligence and projective tests (Mercer & Gibson, 1950; Orne, 1951), physiological responses (Gidro-Frank & Bowersbuch, 1948; Kupper, 1945), and numerous other behaviors reviewed by Barber (1962), Gebhard (1961), and Yates (1961).

According to Barber, the research generated by the concept of age regression may be questioned on methodological grounds (Barber, 1967). The primary criticisms are poor controls and unconvincing dependent variables. Reviews by Gebhard (1961) and Yates (1961), while less critical than Barber, concur in the need for rigor in experimental methodology if the researcher is to establish the validity of age regression.

¹ This study was supported by Grants MH 08061 and HD 00151 from the National Institute of Mental Health. A preliminary report of this study appeared in *Science*, 1968, 159, 1375-1376.

² Requests for reprints should be sent to Richard M. Lundy, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

An approach to the description and understanding of hypnosis is to identify those responses elicited under hypnosis which cannot easily be made under other kinds of suggestions. If responses can be identified which cannot be made at all under other kinds of suggestions, then in the description of those responses may be found the key to understanding the nature of hypnosis. Specifically, the validity of age regression can be better established when responses typical of children but not of adults are produced under age regression and when these same responses are not produced under a waking suggestion. Of course, the possibility of the validity of age regression is increased if those behaviors at the regressed "age" are in fact similar to the actual responses given by children at that age.

It was recently determined that the Ponzo (Figure 1) and Poggendorff (Figure 2) illusions both vary systematically as a function of age (Leibowitz & Gwozdecki, 1967; Leibowitz & Judisch, 1967). The magnitude of the Ponzo illusion increases with age from childhood through adolescence, approaches an asymptote, and decreases again in old age. The Poggendorff illusion decreases in magnitude with age from 5 to approximately 10 yr., and then remains stable.

The *S*'s responses to these illusions could serve as an adequate indication of the facilitation of retrieval of age-appropriate perceptual responses while age regressed. This would not be a matter of memory, since it is highly

unlikely that most Ss were exposed to the specific experimental stimuli at the ages of 5 and 9. In addition, this task seemed especially well suited inasmuch as the age functions of these particular illusions develop in opposite directions. It is virtually certain that Ss had no previous knowledge of these normative functions. Therefore, the possibility of acquiescence to expectations is effectively eliminated. The purpose of the present study was to determine whether retrieval of childhood responses to the Ponzo and Poggendorff illusions would be facilitated by age regression by means of a standard-induction technique.

METHOD

Subjects. Undergraduate students from the introductory psychology courses at the Pennsylvania State University, who answered positively to the statement "I would be willing to participate in a hypnosis experiment" were subjected to a standard-hypnotic-induction procedure and then tested on Barber's Suggestibility Scale (BSS, Barber & Glass, 1962). The first 10 students obtaining a score of 8 on the scale were selected for the present experiment. There were 4 females and 6 males. Further selection of 2 female and 3 male Ss for a control group was determined by the same procedure. The Ss ranged in age from 18 to 21 with a mean of 19.1 yr., and had normal vision either with or without correction.

Apparatus. The apparatus and procedure used in this experiment were identical to that used by Leibowitz and Judisch (1967) for the Ponzo illusion and by Leibowitz and Gwozdecki (1967) for the Poggendorff illusion. For the Ponzo illusion, the stimuli consisted of a series of 21 cards similar to the insert in Figure 1, but with the line nearer to the point of convergence varying in length from $2\frac{1}{2}$ to 5 in. in $\frac{1}{2}$ -in. steps. The other vertical line was constant in length at 4 in. The Poggendorff figure was made of metal with the right half of the diagonal adjustable (see the insert in Figure 2). The center of the figure was 13×4 in. and the diagonal bars were each $7\frac{1}{2} \times 3\frac{1}{2}$ in. with a 45° angle formed between the center bar and the diagonal bar. Details of the construction of the stimuli are discussed in the original publications.

Procedure. Each S was presented the stimuli for the Ponzo and Poggendorff illusions under four conditions: no hypnosis, hypnosis-no regression, hypnosis and regression to age 9, and hypnosis and regression to age 5. Only one condition was presented on a given day, at weekly intervals, as determined by random order. In each hypnosis session a standard-induction procedure and the BSS were administered and then the stimuli were presented. The illusion presented first on a given day was randomly determined.

The Ponzo stimuli were presented in a random order, one card at a time, horizontally on a table

before S who sat approximately 18 in. from the card. The S was given the following instructions: "I am going to show you some cards with lines on them. I want you to point to, or tell me, which of the two vertical lines is longer or bigger." The procedure was adapted from the Leibowitz and Judisch study (1967) which was designed for administration with children. The age-regressed Ss were able to follow these instructions without difficulty. After presentation of the first and second cards in the series E gave no further instructions and simply recorded S's responses without comment. Inasmuch as the order of the cards was random and E was uninformed as to the method for determining the equality value, he was naive regarding the significance of S's responses for any given card at the time of presentation.³ The equality value was determined later by interpolation from the transition point at which S's response changed.

The S was given six trials each session with the Poggendorff apparatus. The right diagonal bar was movable and was set, in an alternating order, significantly out of line above or below the nonmovable, standard left bar. Instructions were given to S to "adjust the right-half of the bar so that it seems to be in a straight line with the left bar." The E demonstrated how the bar could be moved and S was asked if he understood the task. When necessary the instructions were repeated. S's line of regard was kept on a vertical plane relative to the center bar of the illusion apparatus.

Between presentation of the Ponzo and Poggendorff illusions, hypnotized Ss were asked to close their eyes and the age-regression suggestion was reemphasized. The entire presentation time for both illusions was approximately 15 min. for the hypnosis conditions.

A control group of five Ss, selected by the same procedures as the experimental group,⁴ were presented the same stimuli under two conditions: (a) routine instructions without any age regression or function suggestion, and (b) task motivation instructions in which they were told "I want you to be five years of age; you are to regress to age five and we will still be able to talk to each other. You can do this for me and I am going to help you do this by asking you some questions." To facilitate the task-motivation instructions, control Ss were then presented the same list of questions given to the age-regressed experimental Ss. These questions were presented with the suggestion to S that "you are now five years old and I want to ask you some questions," and included inquiries about where he lived, places he visited, his birthday party, and his attitudes toward school and playmates. The S was also asked

³ Even though E was naive regarding the method for determining the equality value and thus did not know the significance of any single response, he did know whether S was age regressed or not. Thus, the possibility of E bias has not been completely eliminated.

⁴ The Ss in both experimental and control conditions were serving in their first hypnosis experiment.

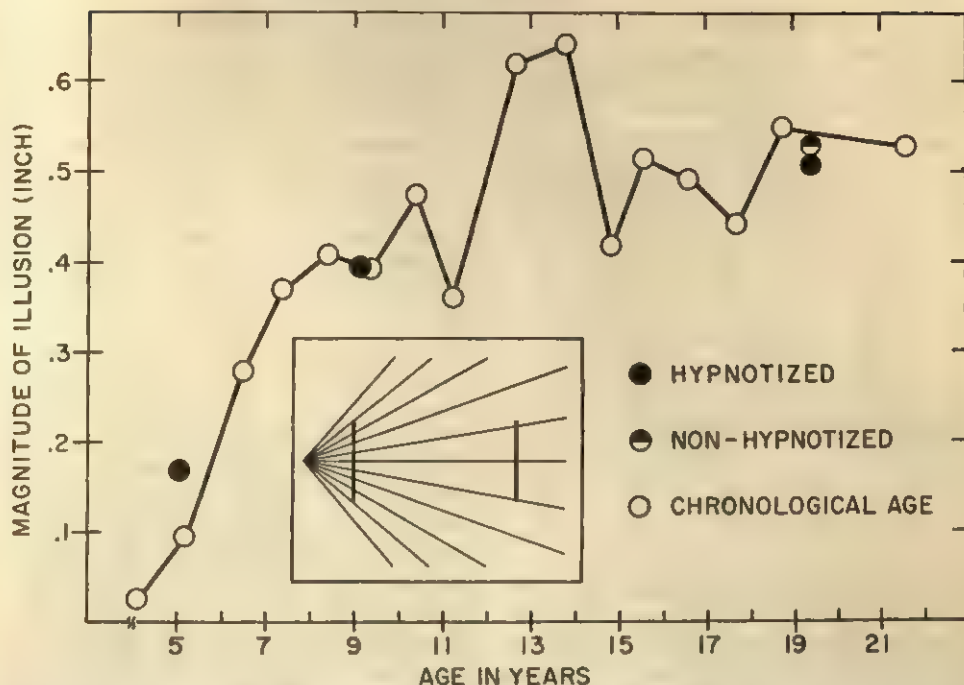


FIG. 1. Ponzo illusion. (The equal vertical lines in the Ponzo illusion (inset) appear unequal. The graph describes the magnitude of this illusion as a function of chronological age, open circles; superimposed are the data obtained under hypnosis, solid circles.) (Reprinted from an article by Michael Parrish, R. M. Lundy, and H. W. Leibowitz published in the March 1968 *Science*. Copyrighted by the American Association for the Advancement of Science, 1968.)

to print his name and to make some numbers. The handwriting samples and the numbers taken from this session and those taken in an earlier S-screening session during which these control Ss were hypnotized and regressed to age 5 were submitted to a faculty member who was asked to determine if the samples, for a given S, were or were not made during the same stage of motor development. The report of the rater that these samples looked like they were made by a "child" during the same period of motor development indicated that Ss would write and make numbers when not hypnotized that were as "child-like" as those when they were hypnotized and regressed. This report, considered with the fact that Ss were willing to answer questions about events during this fifth year, seems to indicate that these control Ss were motivated to cooperate in trying to "be 5 years old." The two control conditions were also presented randomly and separated by 1-wk. intervals.

RESULTS

Ponzo illusion. The magnitude of the illusion was calculated as the difference between the physical equality setting and the subjective equality value interpolated from S's responses. Since the comparison line varied in

$\frac{1}{8}$ -in. steps the interpolation was to the nearest $\frac{1}{16}$ in. The mean scores and standard deviations for each condition of the present experiment, together with the mean scores for the appropriate age group in the Leibowitz and Judisch (1967) study, are presented in Table 1. To facilitate comparison with the normative data obtained as a function of chronological age (CA), the present results are superimposed on the function obtained in the previous experiment (Figure 1). It will be observed that the data in the absence of age regression, either with or without hypnosis, are similar to those obtained with Ss of comparable CA. With regression to age 9, there is a decrease in the magnitude of the illusion which is identical to that obtained with variation of CA. With further regression to age 5, there is a further decrease in the size of the illusion which is not as marked as that obtained with chronological variations.

An analysis of variance leads to a rejection of the hypothesis that the means for the four conditions are equal ($F = 7.27$, $p < .01$). A

TABLE 1

MEAN MAGNITUDE OF PONZO ILLUSION IN INCHES

Data	Conditions			
	No hypnosis	Hypnosis no regression	Hypnosis regressed to 9	Hypnosis regressed to 5
Present				
<i>M</i>	.53	.51	.39	.17
<i>SD</i>	.11	.16	.13	.16
Mean age of groups				
	19.3	19.3	9.2	5.2
Leibowitz and Judisch				
<i>M</i>	.51	.51	.32	.10
<i>SD</i>	.24	.24	.17	.22

Newman-Keuls procedure (Winer, 1962) was employed to determine which pairs of treatment means differed significantly. Table 2 reveals that, with the exception of the differences between conditions no hypnosis and hypnosis-no regression, all mean differences are significant. This effect and the observed correspondence between the present means and the normative means (Figure 1) suggest

TABLE 2

MATRIX OF DIFFERENCE BETWEEN MEANS OF THE FOUR CONDITIONS FOR THE PONZO ILLUSION

Condition	Hypnosis regressed to 5	Hypnosis regressed to 9	Hypnosis	No hypnosis
Hypnosis regressed to 5		22*	34*	36*
Hypnosis regressed to 9			12*	14*
Hypnosis				2
No-hypnosis				

* $p < .05$.

strongly that hypnotic induction facilitated the retrieval of the age-appropriate perceptual response to this illusion. The data for the control group obtained under the condition of no task-motivation were also similar to those obtained by Ss of the same CA, being .68 in. Under the task-motivation instructions, the value of the illusion increased to a mean value of .98 in. This value is in the opposite direction from the value obtained for children of age 5 and suggests that the task-motivation condition did not facilitate the retrieval of age-appropriate perceptual responses to this illusion.

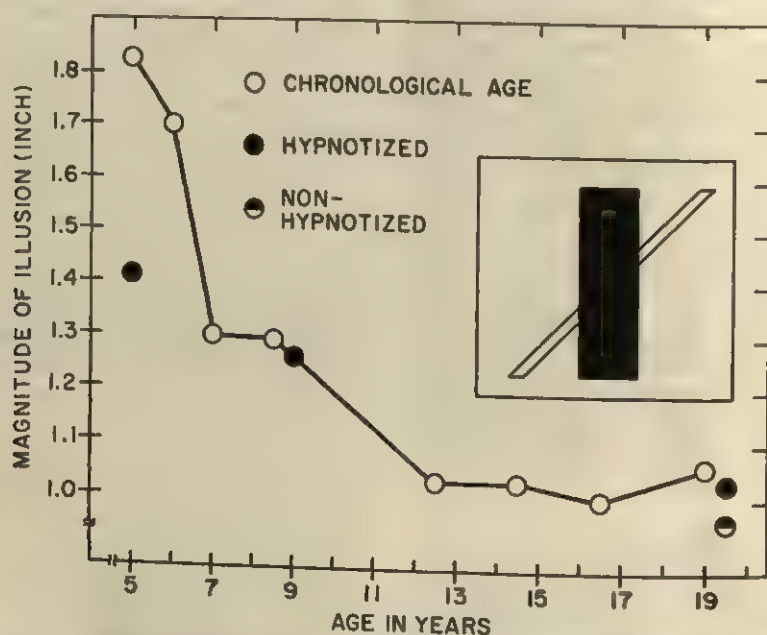


FIG. 2. Poggendorff illusion. (The diagonal line in the Poggendorff illusion (inset) is continuous but appears discontinuous. The graph describes the magnitude of the illusion as a function of chronological age, open circles; superimposed are the data obtained under hypnosis.)

Poggendorff illusion. The magnitude of the illusion was calculated as the differences between the physical quality setting and subjective equality value given by S. A scale, calibrated in $\frac{1}{8}$ -in. steps was attached to the back of the center bar. The means for the six trials under each condition are presented in Table 3 and superimposed on the normative data obtained as a function of CA in Figure 2. The magnitudes of the illusion in the absence of age regression and for the 9-yr. age-regression group are essentially identical to those previously obtained, while the data for the 5-yr. group are in the same direction but not as marked as in the previous study.

An analysis of variance leads to a rejection of the hypothesis that the means of the four conditions are all equal ($F = 4.10$, $p < .05$). A Newman-Keuls procedure computed to determine which pairs of the treatment means differed significantly revealed significant differences between the no-hypnosis and hypnosis-regressed-to-age-5 conditions, and between the hypnosis and hypnosis-regressed-to-age-5 conditions (Table 4). The data for the control group obtained under the no-task-motivation condition were similar to those obtained by Ss of the same CA, being .94 in. However, for the task-motivation condition the mean value was 1.27 in. which corresponds to the hypnosis-regressed-to-age-9 con-

TABLE 4
MATRIX OF DIFFERENCE BETWEEN MEANS OF THE FOUR
CONDITIONS FOR THE POGGENDORFF ILLUSION

Condition	Hypnosis regressed to 5	Hypnosis regressed to 9	Hypnosis	No hypnosis
Hypnosis re- gressed to 5		15	35*	45*
Hypnosis re- gressed to 9			23	30
Hypnosis				7
No-hypnosis				

* $p < .05$.

dition rather than to the hypnosis-regressed-to-age-5 condition.

DISCUSSION

Reponses to the Ponzo illusion appropriate for children were emitted by young adults while age-regressed. The conclusion that age regression facilitates the use or nonuse of visual cues in a manner appropriate to earlier stages of perceptual development is supported by the correspondence between the present data and the normative data based on CA variation. Further support is provided by the fact that Ss given task-motivation instructions were unable to respond in a manner appropriate to 5-yr.-old children. These Ss were selected in the same manner as the experimental groups and were presumably as suggestible as the experimental Ss. Careful postsession interviews indicated that these Ss were surprised to learn that their responses for the task-motivation task were considerably different from those actually obtained from children. Handwriting samples taken while Ss were hypnotically regressed to age 5 and while under task-motivation instructions to write as a 5-yr.-old were not noticeably different. This is of interest in that it suggests that Ss did attempt to follow the task-motivation instructions. However, further investigation regarding this condition should be undertaken. It is difficult to assess the perceived differences in the regression instructions between Ss who have just been given the standard-hypnotic-induction technique and Ss who have had no induction instructions, even when the regression instructions are held constant. It would be of interest to

TABLE 3

MEAN MAGNITUDE OF THE POGGENDORFF ILLUSION
IN INCHES

Data	Conditions			
	No hypnosis	Hypnosis	Hypnosis regressed to 9	Hypnosis regressed to 5
Present				
M	.96	1.03	1.26	1.41
SD	.39	.32	.38	.24
Mean age of groups				
	19	19	8.5	5
Leibowitz and Gwozdecki				
M	1.06	1.06	1.29	1.83
SD	.38	.38	.36	.36

investigate the parameters of the age-regression instructions independently of the standard-hypnotic-induction technique by encouraging Ss to "really feel like you are five years old" within the task-motivation condition.

The retrieval of perceptual responses for the Poggendorff illusion was not as marked as that seen for the Ponzo illusion. The mean obtained for the hypnosis-regressed-to-age-5 group was considerably lower than the normative data for the 5-yr-olds. While it is evident that the mean for the hypnosis-regressed-to-age-9 condition was similar to the normative data for this group, the Newman-Kuels procedure revealed no significant differences between this condition and the hypnosis-regressed-to-age-5 condition. These results, not seen with the Ponzo illusion, suggest that the induction procedure did not facilitate the retrieval of the perceptual cues for this illusion for age 5 to the same degree that it did for the Ponzo illusion.

A satisfactory explanation of this difference is not obvious. However, it should be pointed out that these illusions are most probably mediated by different mechanisms. For example, it has been hypothesized that the Ponzo illusion is a result of inappropriate constancy scaling and is related directly to the individual's environmental experiences (Fleischner, 1898; Gregory, 1963; Tausch, 1954; Von Holst, 1957). A similar explanation does not appear to be appropriate for the Poggendorff illusion.

More importantly, whatever processes of perceptual development are responsible for the observed changes in the magnitude of these illusions as a function of CA, they are apparently not completely irreversible. These results further suggest the possible use of hypnotic age regression as a methodological approach to the study of developmental changes in perception.

REFERENCES

- BARBER, T. X. Hypnotic age regression: A critical review. *Psychosomatic Medicine*, 1962, 24, 286-299.
- BARBER, T. X. "Hypnotic" phenomena: A critique of experimental methods. In J. E. Gordon (Ed.), *Handbook of clinical and experimental hypnosis*. New York: Macmillan, 1967.
- BARBER, T. X., & GLASS, C. B. Significant factors in hypnotic behavior. *Journal of Abnormal and Social Psychology*, 1962, 64, 222-228.
- FILEHNE, W. Die geometrisch-optischen Täuschungen als nachwirkungen der im Körperlichen Sehen erworbenen Erfahrung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1898, 17, 15-61.
- GEBHARD, J. W. Hypnotic age-regression: A review. *American Journal of Clinical Hypnosis*, 1961, 3, 139-168.
- GIDRO-FRANK, L., & BOWERSBUCH, M. K. A study of the planter response in hypnotic age regression. *Journal of Nervous and Mental Disease*, 1948, 107, 443-458.
- GREGORY, R. L. Distortion of visual space as inappropriate constancy scaling. *Nature*, 1963, 199, 678-680.
- HOLST, E. von. Aktive Leistungen der menschlichen Gesichtswahrnehmung. *Studium Generale*, 1957, 10, 231-243.
- KUPPER, H. I. Psychic concomitants in wartime injuries. *Psychosomatic Medicine*, 1945, 7, 15-21.
- LEIBOWITZ, H. W., & GWODZICKI, J. The magnitude of the Poggendorff illusion as a function of age. *Child Development*, 1967, 38, 573-580.
- LEIBOWITZ, H. W., & JUDISCH, J. M. The relation between age and the magnitude of the Ponzo illusion. *American Journal of Psychology*, 1967, 80, 105-109.
- MERCER, M., & GIBSON, R. W. Rorschach content in hypnosis: Chronological age regression. *Journal of Clinical Psychology*, 1950, 6, 352-358.
- ORNE, M. T. The mechanisms of hypnotic age regression: An experimental study. *Journal of Abnormal and Social Psychology*, 1951, 46, 213-225.
- TAUSCH, R. Optische Täuschungen als artifizielle Effekte der Gestaltungsprozesse von Grossen- und Formenkonstanz in der natürlichen Raumwahrnehmung. *Psychologische Forschung*, 1954, 24-25, 299-348.
- TRUE, R. M. Experimental control in hypnotic age regressed states. *Science*, 1949, 110, 583-584.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- YATES, A. J. Hypnotic age regression. *Psychological Bulletin*, 1961, 58, 429-440.

(Received September 13, 1968)

LEARNING AND USE OF LOGICAL SYMBOLS BY DEAF AND HEARING SUBJECTS¹

SOLIS L. KATES²

University of Massachusetts

Deaf Ss were as able as hearing Ss matched on age, IQ, and socioeconomic background in learning and using logical concepts, and superior to hearing Ss matched on written language comprehension, IQ, and socioeconomic background in learning these concepts. Verbal language comprehension apparently does not possess an exclusive determining influence in learning and grasping the significance of these concepts. It seems justifiable that the ability to identify correspondences between logical symbolic expressions and their stimulus combinations, to verbalize accurately the rules of these correspondences, and to use learned symbols in new contexts constitute increasingly more difficult levels of conceptual behavior. Supplementary verbalization of rules leads to more successful transfer than learning the rules without correct verbal formulations.

Are deaf adolescents able to match logical symbols with their appropriate referents in a series of corrected trials? Are they able to describe the significance of these symbols correctly after these acquisition trials? And are they able to employ these symbols in a test situation with new stimulus features? These questions are explored in this study, which was essentially a replication of an investigation by Furth and Youniss (1965).

The logical symbols, \cdot , $/$, and $\bar{\cdot}$, represent three different connectives—"and," "or," and "not," respectively. These symbols were combined with the letters, B and R—B for blue and R for red—to form three different symbolic expressions representing three different concepts—the conjunctive ($B \cdot R$), the exclusive-disjunctive (B/R), and the conjunctive-absence ($\bar{B} \cdot \bar{R}$). In the acquisition phase, where corrections were given, each trial required the matching of certain color combi-

nations with one of the three expressions incorporating the letters B and R and the logical symbol representing the connective. In the test phase, these same logical symbols, standing for the same connectives, were used but with new letters, C and D, referring to combinations of the new stimulus features, circular and dark.

Furth and Youniss (1965) have identified two different modes of learning the relationships between the symbolic expressions, consisting of the logical connectives and the letters B and R, and the corresponding stimulus combinations. In the first, learning this connection is probably a result of remembering their perceptual correspondence. In the second, the connection between the symbolic expression and the stimulus combination is learned, but this learned connection is or can be divorced from the concrete data—that is, from the stimulus features on which it was originally built. Such learning approximates the formal operations that Piaget (1957) has described. When learned in the latter way, this connection may be successfully applied to new stimulus combinations when it is relevant. In order to evaluate whether this second mode of learning has occurred, the successful Ss from the acquisition phase were examined on the test phase.

Deaf Ss tested by Furth and Youniss (1965) could not apply these logical symbols to new stimulus combinations. Only 7% of their deaf Ss, aged 12 through 20 yr., who at-

¹ This research was supported, in part, by Research Grant No. HD 00336 from the United States Public Health Service. This paper was presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 20, 1968. The author expresses appreciation to Anne Young and Karen Bruscoe for their assistance in conducting this research. The author is also grateful to George Pratt and the supervising teachers at the Clarke School for the Deaf for their cooperation, and to the public schools in Northampton and Easthampton, Massachusetts.

² Requests for reprints should be sent to the author, Psychology Department, University of Massachusetts, Amherst, Massachusetts 01002.

tained criterion on the acquisition task were successful in transfer. However, 64% of the urban, hearing Ss aged 12 through 14 yr. successfully matched these logical connectives with the new stimulus combinations in the test phase. Furth and Youniss ascribed the poor transfer ability to the deaf Ss' failure to look spontaneously for an underlying basis linking logical connectives with their appropriate referents. Satisfied from previous research that deaf Ss are able to comprehend and use these logical symbols when specifically taught, Furth (1966) believed that they learned by rote, with a lack of intellectual curiosity in tasks where comparable hearing Ss attempt to discover the significance of these symbols. Furth believes that such impairment of spontaneous learning in a discovery task is related, in part, to the restricted experience of the deaf Ss and not simply to limited verbal language.

Furth and Youniss used deaf Ss who were unable to comprehend written language at the difficulty level of ordinary conversation; possibly this language factor may have played some role in their findings. Are deaf Ss who can understand written connected language at some minimal level, such as the fifth-grade level, able to learn and then employ the learned logical symbols in new situations? It was thought that deaf Ss, matched with hearing Ss on the relevant variables, would be as able to learn and then employ the learned logical symbols in new stimulus contexts, provided that they had this minimal language comprehension. The enrichment of their experiences which may result, in part, from their language ability should help the deaf in learning the connections between symbolic expressions and stimulus combinations. In addition, the support given by a correct verbal formulation should be of considerable value in using the logical symbol in a new context (Inhelder & Piaget, 1958, p. 253).

Consequently, the author obtained verbal statements describing the significance of the symbolic expressions following the completion of the acquisition phase. It is believed that, for both deaf and hearing Ss, attaining criterion on the acquisition task and giving correct verbal descriptions of the symbolic ex-

pressions would be associated with greater success on the test phase. These verbal formulations, when present with the learned connections, should help direct the responses of Ss to relevant stimulus combinations and should help resist the interference of irrelevant stimulus features. One can say that there is a more "abstract" attitude, in the Goldstein sense, when the learned connection and its supplementary verbal formulation are present than for the learned connection alone.

It is possible that this study may have some relevance to the contribution of age and language to the learning of logical symbols. When deaf Ss are matched on written language comprehension with hearing Ss, they are usually older and may show the impact of more contacts with the physical and social environment on their learning of logical symbols. If the deaf are more able than these language matched, but younger, hearing Ss to acquire the connection between the logical symbols and their respective instances, and to transfer this learning to a new task, it may provide evidence for the role of age-related adaptations on the development of logical thinking. It would seem to show that mental operations basic to logical thinking may be refined or extended by language but are not dependent on language for their emergence. For example, in discussing the logical operation of inclusion, Piaget (1957) stated that "the child has not simply to carry out a verbal or symbolic translation of the perceptual data, but an operational composition or decomposition of its elements . . . [p. 5]." Furthermore, Piaget (1957, p. 6) continued on to state that such perception involves knowledge of logical operations, which require age-related contacts with the environment. But if the deaf and language matched hearing Ss are not different in their learning of these logical relations, the effect of language on logical behavior may have to be given somewhat additional consideration.

The conjunctive concept has been described as easier than the exclusive-disjunctive or conjunctive-absence concepts for adolescents (Youniss & Furth, 1964). Yet Neisser and Weene (1962) have indicated that the conjunctive-conjunctive-absence, and exclusion (similar to the exclusive-disjunctive concept)

concepts are not significantly different for college students. This question can be further examined in the context of this investigation.

METHOD

Subjects

Sixteen deaf *Ss* were selected from the Clarke School for the Deaf in Northampton, Massachusetts, who were between the ages of 11 and 17 yr., who had been deaf from birth or before the age of 3, and could not have learned connected English language without special instruction. There were 8 girls and 8 boys.

A group of 16 hearing *Ss* were selected from the Kennedy Junior High School in Northampton, Massachusetts, and from the Northampton High School in Northampton, Massachusetts. They were between 11 and 17 yr. of age and were matched with the deaf *Ss* on chronological age, sex, socioeconomic level, and IQ (hereafter known as hearing-age *Ss*). Another 16 hearing *Ss* between the ages of 9 and 11 yr. were selected from the fourth, fifth, and sixth grades of the Easthampton, Massachusetts, elementary school system; they were matched with the deaf *Ss* on IQ, sex, socioeconomic background, and comprehension of written language (hereafter known as hearing-achievement *Ss*). The IQ scores were obtained by using the Picture Arrangement, Block Design, and Coding subtests of the appropriate Wechsler Intelligence Scales for children. The average of the word and paragraph meaning scores on the Stanford Achievement Test were used to match hearing and deaf *Ss* on written language comprehension. The Hollingshead scale (1957) was used to provide a numerical rating of the fathers' occupations for matching groups on socioeconomic level. Table 1 presents the means and standard deviations for *Ss* on these independent and control variables.

Stimulus Materials

Phase 1: Acquisition. In the acquisition task 48 stimulus cards, each measuring 4 × 6 in., were used. Sixteen stimulus cards exemplified the conjunctive concept, 16 cards the exclusive-disjunctive concept, and 16 cards the conjunctive-absence concept. Each of the 16 conjunctive concept cards was colored half red and half blue (with the left-right color position counterbalanced). The correct response for the conjunctive card was *B·R*, representing a conjunction of blue and red.

The 16 exclusive-disjunctive concept cards consisted of 4 red and yellow cards, 4 red and green cards, 4 blue and yellow cards, and 4 blue and green cards. (Again the position of these colors was counterbalanced.) The disjunctive response of *B/R* was to be given to any card containing blue and another color besides red, or containing red and another color besides blue.

Each of the 16 conjunctive-absence concept cards was colored yellow and green. (Again the position

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR AGE, IQ, WRITTEN LANGUAGE COMPREHENSION, AND SOCIOECONOMIC BACKGROUND

Group	Age	IQ	Stanford reading achievement ^a	SE ^b
Deaf				
\bar{X}	15-0	109.3	5.3	3.8
<i>SD</i>	27 mo.	10.0	.8	1.6
Hearing age				
\bar{X}	15-1	109.9	—	4.4
<i>SD</i>	25 mo.	9.5		1.3
Hearing achievement				
\bar{X}	10-11	109.9	5.3	4.3
<i>SD</i>	10 mo.	14.9	.9	1.3

Note.—*N* = 16.

^a In grade level.

^b Hollingshead Socioeconomic Index.

of these colors was counterbalanced.) The correct response for this class was *B·R*, which was to be given to any card which was neither blue nor red.

Phase 2: Test. Here 18 cards were used, with 5 cards illustrating the conjunctive concept, 5 cards the conjunctive-absence concept, and 8 cards the exclusive-disjunctive concept. In the test task, there was a change in the two critical elements; instead of blue and red, the two elements were now circular form and dark color. The 5 conjunctive cards contained dark colored circular figures (circular and dark, *C·D*); the 8 disjunctive cards included 4 cards having dark colored angular figures and 4 cards with light colored circular figures (either circular or dark, *C/D*); and the conjunctive-absence cards contained light colored angular figures (neither circular nor dark, *C·D̄*).

Procedure

Phase 1: Acquisition. Each *S* was tested individually in both the acquisition phase and the test phase. In the acquisition phase *S* was instructed to look carefully at colored cards shown one at a time, and to choose which of the three answers (*B·R*, *B/R*, *B·R̄*) he thought belonged with that colored card. He was told that in the beginning he would have to guess, but that corrections given by *E* would help him find the correct answer for each card. He was not allowed to make any marks or changes in previous answers. There were a total of 48 trials, or eight blocks of 6 trials each; within each block, there were two instances of each concept shown in random order. After the 48 trials, *S* was asked to write the meaning of each of the three logical symbolic expressions.

Phase 2: Test. The test phase consisted of 18 stimulus instances with circular and dark as the critical elements presented in random order. The *S* was again instructed to look carefully at the cards and, using what he had learned before (in the

TABLE 2

CHI-SQUARE VALUES AND FISHER ϕ VALUES FOR COMPARISONS OF THE NUMBER OF SUBJECTS WHO ATTAINED CRITERION IN THE ACQUISITION PHASE

Condition	Number who attained	Number who did not attain	Statistic
Deaf	16	0	$\chi^2 = 13.9, p < .001$
Hearing achievement	5	11	ϕ (Fisher) = .0001
Deaf	16	0	ns
Hearing age	13	3	
Hearing age	13	3	$\chi^2 = 6.22, p < .02$
Hearing achievement	5	11	ϕ (Fisher) = .005

acquisition task), decide which answer (now C·D, C/D, $\bar{C}\cdot\bar{D}$) belonged with each card. The test phase was administered in the same way as the acquisition phase except that no corrections were given, and *S* circled the correct answer instead of writing it in a blank. Immediately after the last trial, *S* was asked to write the meaning of the three response symbols. Throughout the testing, *E* was careful not to verbalize the three symbols or their meanings.

Each *S* was considered to have attained criterion on the acquisition task if he fulfilled either of the following two standards: (a) he had to achieve 10 consecutively correct responses with no more than one error following these 10 responses or (b) he had to respond correctly in the last seven trials. On the test task, *S* was considered to have attained criterion if he fulfilled one of the following two conditions: (a) he had to have 10 consecutively correct responses with no more than a total of 3 wrong responses and no more than 1 wrong response after the 10 consecutively correct responses or (b) he had to have the last 8 (or more) consecutively correct with no more than a total of 3 wrong responses.

Finally, in order to evaluate the written descriptions given to each of the three symbolic expressions, a scoring system was devised with illustrative examples for assigning a score of 3, 2, or 1 to each description. Essentially, the description of a concept received a score of 3 if it included both the appropriate critical elements and the proper logical connective or its equivalent. Examples are blue and red for B·R, circular form and dark color for C·D, circular or dark for C/D, and neither blue nor red for $\bar{B}\cdot\bar{R}$. A score of 2 was given to descriptions that mentioned the two critical elements without specifying the proper logical connective, such as blue-red for B·R, a circle, dark color for C·D, blue with yellow for B/R, green and yellow for $\bar{B}\cdot\bar{R}$, and a pink triangle for $\bar{C}\cdot\bar{D}$. A score of 1 was given when the proper critical elements were not specified and the appropriate logical connective was not included,

such as blue for B·R, circle for C/D, and triangle for $\bar{C}\cdot\bar{D}$. A maximum score of 9 was possible for each phase. Scoring was done independently by two judges, reliability was .98. All protocols were scored without knowledge of *S*'s identity or group.

RESULTS

The data indicate that deaf *S*s and hearing-age *S*s were not significantly different in the number who attained criterion in the acquisition phase (Table 2). However, the number of hearing-achievement *S*s who attained criterion in the acquisition phase was significantly smaller than either the number of deaf *S*s or the number of hearing-age *S*s (Table 2).

When the data for each *S* group of 16 were analyzed for the correct verbal description of the three concepts after the acquisition phase (Table 3), 7 of the deaf *S*s, 6 of the hearing-age group, and none of the hearing-achievement group were able to describe the three concepts correctly.

Further, of those 16 deaf *S*s who attained criterion in the acquisition phase, just 5 *S*s were able to attain criterion in the test phase. Similarly, 4 of the 13 hearing-age *S*s who attained criterion in the acquisition phase were able to achieve criterion in the test phase. None of the 5 hearing-achievement *S*s who attained criterion in the acquisition phase were able to succeed in reaching criterion in the test phase. Statistical tests evaluating the differences in the proportion of successful *S*s in each of the three groups who reached

TABLE 3

CHI-SQUARE AND FISHER ϕ VALUES FOR COMPARISONS OF THE NUMBER OF SUBJECTS WHO VERBALIZED CORRECTLY THE THREE CONCEPTS FOLLOWING ACQUISITION

Condition	Number who verbalized correctly the three concepts	Number who did not verbalize correctly the three concepts	Statistic
Deaf	7	9	$\chi^2 = 6.6, p < .02$
Hearing achievement	0	16	ϕ (Fisher) = .003
Deaf	7	9	ns
Hearing age	6	10	
Hearing age	6	10	$\chi^2 = 5.1, p < .05$
Hearing achievement	0	16	ϕ (Fisher) = .007

criterion in the test phase did not produce any significant findings (Table 4). It should be recalled that about one-third of the Ss in the deaf and in the hearing-age groups who were successful in the acquisition phase were also successful in the test phase.

The author also examined how correct verbalizations together with achieving criterion on the acquisition phase affected performance in the test phase. Table 5 shows the distribution of Ss who attained criterion on the acquisition phase together with correct descriptions of the three concepts, and of Ss who succeeded in attaining criterion on the acquisition phase but were not able to describe the three concepts correctly. Chi-square and Fisher probability tests show that there was a significant finding indicating that correct verbalization and successful performance in the acquisition phase was associated with significantly greater numbers of Ss who attained criterion on the test phase. A significant relationship was also obtained for the hearing-age Ss but there was only a trend (at the .10 level) for the deaf Ss (Table 5).

There were no significant differences in difficulty of the three concepts for the deaf and for the hearing-achievement Ss on the acquisition phase. However, on the acquisition phase the conjunctive concept was significantly easier for the hearing-age Ss than the exclusive-disjunctive and conjunctive-absence concepts whereas the latter two were not significantly different. The F ($df = 2/45$) value for the hearing-age group was 5.9, significant at the .01 level. Also, the percentage of correct trials for those Ss in each of the three groups

TABLE 4
CHI-SQUARE AND FISHER p VALUES FOR COMPARISONS
OF THE NUMBER OF SUBJECTS WHO WERE
SUCCESSFUL ON ACQUISITION AND
WHO ATTAINED CRITERION IN
THE TEST PHASE

Condition	Number who attained criterion in test phase	Number who did not attain criterion in test phase	p
Deaf	5	11	ns
Hearing achievement	0	5	
Deaf	5	11	ns
Hearing age	4	9	
Hearing age	4	9	ns
Hearing achievement	0	5	

who had been unable to attain criterion on the test phase although they had been successful on the acquisition phase was compared. Analyses of variance showed no significant differences in the difficulty of the three concept types on the test phase for the three S groups for those Ss who succeeded on the acquisition phase but failed the test phase.

DISCUSSION

The deaf Ss were as able as the hearing-age Ss to attain criterion in the acquisition phase. Inasmuch as the acquisition phase involved establishing the relationships between stimulus materials which were largely perceptual in nature, through an immediate corrective feedback procedure, it could be anticipated that the deaf Ss would do as well as the hearing-age Ss. There was a significantly greater number of deaf and hearing-age Ss

TABLE 5
RELATION BETWEEN ACQUISITION PHASE AND CORRECT VERBALIZATION AND
ATTAINING CRITERION ON TEST PHASE WITH CORRESPONDING
CHI-SQUARE OR FISHER " p " VALUES

Group	Ss who attained criterion in test phase	Ss who did not attain criterion in test phase	Statistic
Ss (deaf plus hearing age) in Group 1	8	5	$\chi^2 = 6.3, p < .02$
Ss (deaf plus hearing age) in Group 2	1	15	p (Fisher) = .001
Deaf Ss in Group 1	4	3	p (Fisher) = .095
Deaf Ss in Group 2	1	8	
Hearing-age Ss in Group 1	4	2	p (Fisher) = .025
Hearing-age Ss in Group 2	0	7	

Note.—Group 1 = Ss who attained criterion in acquisition phase plus verbalizing correctly the three concepts; Group 2 = Ss who attained criterion in acquisition phase but did not verbalize correctly the three concepts.

than hearing-achievement Ss who attained criterion in the acquisition phase. This greater success of the deaf Ss cannot be attributed to the comprehension of written language inasmuch as they were matched on this variable with the hearing-achievement group. Verbal language comprehension, it seems, was not the crucial factor mediating the learning of the correspondence between symbolic expressions and their stimulus combinations.

When evaluated on the ability to give correct verbal descriptions of the three logical symbolic expressions, the deaf and the hearing-age Ss proved to be significantly superior to hearing-achievement Ss. Even in the case of verbal descriptions of the logical symbolic expressions, similar ability to comprehend connected written verbal language apparently plays a small role. If the grasp of the significance of the logical symbolic expressions were solely dependent on similar comprehension of written verbal language, then deaf Ss and hearing Ss matched on this variable should not have been significantly different. But inasmuch as they were significantly different on verbalization, then again it may be inferred that language ability was not the determining influence in understanding the three logical symbolic expressions.

Consequently, on the basis of this finding and the one immediately described above, age seems to be more vital than written language capacity for the appropriate acquisition and verbal description of logical concepts. Probably the effects of age include greater experience with these concepts in daily life and with the component operations necessary to the understanding of these concepts, such as combining, dissociating, ordering, and setting up of correspondences. As Piaget (1957) has stated, capacity for logical thought depends more upon knowledge of the many necessary component operations than upon verbal symbolic skills. If one assumes that greater age is related to greater experience with and knowledge of these logical operations, then it appears that these results offer partial support to the aforementioned formulation of Piaget. The present results emphasize that written language knowledge cannot substitute for greater experience related to increased age, particularly at the age levels

studied here. The conclusion is that the development of skill in certain mental operations must precede the ability to describe verbally the logical symbolic expressions correctly; and skill in the use of these mental operations probably emerges from the active age-related adaptation efforts of the growing individual in his attempts to cope with environmental fluctuations.

The successful acquisition of the connection between logical symbolic expressions and their stimulus combinations affords a small guarantee that the learned symbols representing the logical connectives will be employed appropriately in a new context. The small percentage of deaf and hearing-age Ss who were able to use these logical symbols appropriately in the new context suggests that for most Ss their acquisition success took the form of perceptual correspondences between logical symbolic expressions and stimulus referents. In the hearing-achievement group, there were no Ss who could make this transfer. While no significant differences in deaf, hearing-age Ss, and hearing-achievement Ss could be established, one finding is nonetheless of interest. The substantial similarity in the transfer behavior of deaf and hearing-age Ss contradicts the Furth and Youniss (1965) result. One possible reason for this discrepancy is that in the present study the deaf and hearing Ss were matched as precisely as possible on the variables of age, IQ, and socioeconomic background whereas Furth and Youniss did not match their deaf and hearing Ss. When deaf and hearing Ss have been matched and are approximately similar on these control variables, they demonstrate an approximately equivalent capacity for spontaneous efforts to discover the significance of the logical symbolic expressions.

When one accepts the correct verbalization of the three concepts as an alternative criterion of learning in the acquisition phase, one notes that there is greater predictability for logical behavior in the new stimulus context of the test phase. But there is a limiting factor even when successful verbal behavior characterizes the learning of the logical symbolic expressions by deaf and hearing-age Ss. Inasmuch as approximately two-thirds of these Ss used these learned symbols success-

fully in the new context, it is likely that the correct verbal descriptions of about one-third of these Ss were not entirely free of the concrete stimulus features of the acquisition phase to which they applied. Some of these correct descriptions probably represented verbal statements that could not be transformed into more abstract rules about the logical symbols applicable to the same logical combinations but composed of different stimulus features. It seems justifiable to infer that the ability to identify correspondences between logical symbolic expressions and their stimulus combinations, the ability to describe accurately the rules of these correspondences, and the ability to use these learned symbols in new contexts constitute increasingly more difficult levels of conceptual behavior related to age and possibly other factors such as intelligence.

Another point of discussion deals with the significant relationship between attaining criterion in the acquisition phase plus correctly describing the logical symbolic expressions and successful behavior in the test phase. When Ss attained criterion in the acquisition phase without correct statements about the logical symbols, they showed significantly less success in employing these symbols in the new context of the test phase. Possibly, this supplementary verbalization of the rule helps by providing Ss with an ordering principle or cognitive structure which supports transfer behavior. When faced with different stimulus combinations that may disturb the employment of the learned symbol, Ss who have the verbal statement of this rule available may possess an additional resource for initiating regulative control over their behavior. When this ordering principle is available in verbal form, there seems less likelihood that changed stimulus referents will disturb appropriate transfer behavior to any significant degree. At this point, the general finding that verbal ability per se is not the determining factor in

the successful learning of logical symbols should be emphasized. However, the ability to put these logical symbols, once they have been learned, into verbal language apparently makes a significant contribution to transfer behavior.

Only the hearing-age group found the exclusive-disjunctive concept more difficult than the other two concepts, and this finding relates specifically to the acquisition phase. On this acquisition phase, the other two groups did not experience any concept as more difficult than the others. On the test phase, where the relative difficulty of any concept or concepts would possibly yield information about which ones prevented attainment in a new context, there were no differences in concept difficulty. One can conclude that these three concepts present about the same degree of difficulty for most adolescent Ss in a learning and correction phase and about the same degree of difficulty for all Ss when they have to deal with these concepts involving new stimulus elements in a phase where no correction is given.

REFERENCES

- FURTH, H. G. *Thinking without language*. New York: Basic Books, 1966.
- FURTH, H. G., & YOUNISS, J. The influence of language and experience on discovery and use of logical symbols. *British Journal of Psychology*, 1965, 56, 381-390.
- HOLLINGSHEAD, A. B. Two factor index of social position. New Haven, Conn.: Author, 1957. (Mimeo)
- INHELDER, B., & PIAGET, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic Books, 1958.
- NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, 64, 640-645.
- PIAGET, J. *Psychology and logic*. New York: Basic Books, 1957.
- YOUNISS, J., & FURTH, H. G. Attainment and transfer of logical connectives in children. *Journal of Educational Psychology*, 1964, 55, 357-361.

(Received September 24, 1968)

PSYCHOPATHOLOGY AND SEMANTIC INTERPRETATION OF AMBIGUOUS WORDS¹

THOMAS B. BENJAMIN² AND NORMAN F. WATT³

Harvard University

Homograph measures of vocabulary and semantic interpretation were used to test several theories of language disorder in schizophrenics and aphasics. Aphasics showed a loss of abstract vocabulary, but this did not appreciably affect their interpretation of word meanings in sentences. The raw performance of the schizophrenics was consistently worse than the controls' and better than the aphasics', despite comparability among the groups in education and verbal intelligence. However, there was no evidence of concretistic "regression" in any of the language measures of the schizophrenics. Schizophrenics were somewhat more inclined than controls to select primary meanings (strongest mediating responses) for ambiguous words in sentences, but this was not attributable to lack of knowledge of secondary meanings. For the most part, the results were consistent with Chapman's theory of primacy bias in schizophrenia.

One reason for the persistent interest in the semantic aspects of language in mental patients, historically, was the early recognition by clinicians that semantic changes take place, especially in schizophrenia and certain disorders involving known damage to the brain. Disturbance of association was regarded by Bleuler (1950) as a primary symptom of schizophrenia "insofar as it involves a diminution or leveling of the number of affinities [p. 350]." On the other hand, Piro (1958) emphasized referential widening and "semantic dispersion" in schizophrenia. Concretism has long been observed in the problem

solving and language of mental patients (Flavell, 1956; Goldstein & Scheerer, 1941; Milgram, 1959). The great interest in studying these disturbances in schizophrenia is partially explained by the prospect of finding measurable organic causes for it that are common to brain disorders (e.g., aphasia) which clearly have less mysterious origins. Thus, empirical studies of the observed disturbances of association and abstraction may have significance for the etiology of schizophrenic symptoms.

Another reason stems from a practical issue in clinical assessment: whether vocabulary knowledge is the intellectual function most resistant to deterioration and, therefore, the most stable index of premorbid functioning (Terman, 1918). Babcock (1930) developed an index of deterioration based on the rationale that deterioration occurs first in areas of new learning and the formation of new associations and last in those areas learned earliest in life (vocabulary). Critics have pointed out that real deterioration in conceptual level or organization can occur without showing an appreciable vocabulary loss because, in contrast to tests of other functions, the standards for acceptable achievement on most vocabulary tests are so low (Feifel, 1949; Harrington & Ehrmann, 1954; Yacorzynski, 1941). While this evidence weighs against Babcock's hypothesis, it does not disprove its validity. Stronger contrary evidence comes from direct

¹ The investigation was supported by an undergraduate research grant from the National Science Foundation and by Grant MH-11951-01 from the National Institute of Mental Health. The data for the schizophrenics and alcoholics were drawn from a senior honors thesis submitted by the first author to the Department of Social Relations, Harvard University, under the second author's sponsorship. The authors are greatly indebted to Lewis Sherman and Harold Goodglass, the Chiefs of Psychology Service at the Brockton Veterans' Administration and Boston Veterans' Administration Hospitals, respectively, for making it possible to test the patients in this study, and to Maxine Schuster and Joy Yeager for extensive assistance in the collection and scoring of data and statistical analysis.

² Now at the law firm of Haussermann, Davison and Shattuck, Boston, Massachusetts.

³ Requests for reprints should be sent to Norman F. Watt, Department of Social Relations, William James Hall, Harvard University, Cambridge, Massachusetts 02138.

tests of vocabulary deficit associated with brain damage (Williams, Lubin, & Gieseeking, 1959) or a schizophrenic episode (Lubin, Gieseeking, & Williams, 1962). Still, the practical need to measure intellectual deterioration remains, and clinicians continue to use vocabulary scores in this way, viz., the comparison of "hold" versus "don't hold" subtests of the Wechsler Adult Intelligence Scale (WAIS).

Finally, several theorists (Goldstein, 1948; Vigotsky, 1934; von Domarus, 1944; Werner, 1948) have expounded a regression principle: that in certain forms of psychopathology the language of adults regresses to a more primitive, concrete level. Brown (1958) showed that this principle is a corollary of a popular developmental "law of linguistic progression," which holds that progression from the concrete to the abstract characterizes the evolution of the species from animal to man, the historical development of contemporary languages and the transformation of children into adults. It has been inferred that if language declines with mental disorder according to the regression principle, then language progresses with normal development according to the general law, showing increasing mastery of abstraction with age. This reasoning assumes an intermediate premise which should itself be tested: that linguistic decline retraces the path of linguistic growth, rather than taking some other course. That obstacle has not dampened the enthusiasm for the inference in the past, and the relevance of language disorder for general linguistic theory remains an open question.

Homograph Interpretation

A homograph is a word that can assume two or more different meanings without any change in spelling. Usually a homograph is likewise a homophone, being pronounced the same regardless of its meaning, but there are exceptions. It is a promising tool for experimental psycholinguistics because it offers a means of studying systematically various semantic and syntactic aspects of written language. Pursuing the line of research on clinical assessment aforementioned, Willner (1965) designed an Unusual Meanings Vocabulary Test, comprised of homographs having a very familiar meaning and an unusual

meaning. Whereas most vocabulary tests measure the number of words of progressively decreasing familiarity for which *S* can give one meaning, his test required *S* to select from five alternatives the uncommon meaning of a familiar homograph. Willner found that in comparison with normal controls, schizophrenics knew significantly fewer unusual meanings than usual meanings (the latter measured by a standard WAIS Vocabulary test). This discrepancy did not reach significance for brain-damaged patients. Both patient groups were significantly impaired in abstract reasoning. From these results Willner speculated that impaired reasoning by schizophrenics may be attributable to pathological alterations in the meanings of words used in the abstraction process, whereas brain damage impairs reasoning by diminishing the ability to form relationships between words without altering their meaning.

The meaning of a homograph in a sentence is ambiguous, and correct (consensual) interpretation depends on syntactic and referential constraints in the sentence, as well as the relative "mediating response strengths" for the alternative meanings possible. Chapman, Chapman, and Miller (1964) found that errors of interpretation in sentence contexts by both normals and schizophrenics were biased toward the primary alternative, that is, the meaning which first comes to mind when the homograph is encountered *out of context*. The primacy bias for the schizophrenics was significantly greater than for the normals. Rather than being qualitatively different, this schizophrenic deviation was merely an exaggeration of a normal language pattern. In a related experiment, the same authors found no differences between the two groups in knowledge of primary and secondary homograph meanings, so that the stronger primacy bias of the schizophrenics could not be attributed to loss of secondary meanings from their vocabulary. This finding opposes the results of Willner's study; however, the format of Willner's vocabulary test was slightly different and his items were considerably more difficult.

The present investigation was designed to test whether schizophrenics or aphasics differ from controls in (a) making more primary

than secondary errors of homograph interpretation in context, (b) making more concrete than abstract errors of homograph interpretation in context, (c) knowing fewer secondary than primary homograph meanings, (d) knowing fewer abstract than concrete homograph meanings, and (e) making more interpretational errors in context attributable to vocabulary deficiency. Chapman's theory predicts that for schizophrenics the first hypothesis would be confirmed, but the third and fifth (as related to primacy bias) would be rejected. Willner's theory predicts confirmation of the third hypothesis for schizophrenics but rejection of it for aphasics. The second and fourth hypotheses are formulations of the popular "principle of linguistic regression" in terms of semantic disambiguation and vocabulary, respectively.

METHOD

Subjects

The experimental Ss were 44 chronic Veterans' Administration patients with a primary diagnosis of schizophrenic reaction and no indications of brain damage, and 11 Veterans' Administration patients with acute brain syndrome who showed primary aphasic symptoms and no indications of psychosis. Almost all of the schizophrenics were diagnosed as chronic undifferentiated or paranoid type. Seven aphasics had suffered cerebral vascular accidents, 3 had received head injuries and the etiology in one case was unknown. The control groups were 22 chronic Veterans' Administration patients with a primary diagnosis of alcohol addiction but no indications of psychosis or obvious brain damage, and 20 normal volunteers who were either state hospital employees or working-class acquaintances of the state hospital personnel officer. The normals were the principal control group. The alcoholics provided a measure of control for chronic hospitalization and drug intake, which might influence the schizophrenics' performance. The schizophrenics and alcoholics had been in and out of hospitals for 15 and 8 yr., respectively, on the average, and 89% of the former and 82% of the latter were receiving tranquilizing drugs. Seven alcoholics were also taking antabuse.

No S was mentally deficient. They participated without special incentives, except that the normals were paid \$2.50. The means and standard deviations for length of education were 11.25, $SD = 2.07$ for the normals, 10.82, $SD = 2.52$ for the alcoholics, 10.47, $SD = 2.65$ for the schizophrenics, and 10.91, $SD = 2.87$ for the aphasics. On the Shipley-Hartford Vocabulary Scale they were 27.95, $SD = 5.69$ for the normals, 26.05, $SD = 9.86$ for the alcoholics, 25.20, $SD = 7.44$ for the schizophrenics, and 24.29, $SD = 9.02$ for the seven aphasics tested. The four groups did not differ in education, parental social class, or

vocabulary scores, although the schizophrenics were slightly lower than the normals on the latter ($p < .20$). The normals were younger on the average (35 yr.) than the alcoholics (42 yr.) and the schizophrenics (41 yr.), and the aphasics (54 yr.) were older than all other groups, but the few associations found between age and language performance scores were explained by their joint correlation with education. As the groups were comparable in length of education, the age differences were considered inconsequential.

Measures

The Shipley-Hartford Vocabulary Scale was used as a measure of current verbal intelligence and "cognitive presence." The interpretation test consisted of 48 sentences, each containing a homograph. For example, one item read as follows:

The policeman took his club to the scene of the riot.

- This means: A. He took a heavy stick.
B. He took a set of keys.
C. He took a group of people.

Later in the test the following item occurred: "Mr. Jones took his son's club on a picnic." The same three alternative interpretations followed, in the order of A, C, B. A heavy stick is the consensually correct answer in the first item but the error alternative in the second. A group of people is the error alternative in the first item but the correct answer in the second. A set of keys is an irrelevant alternative in both items, included to measure random responses.

The homographs were selected from a list of over 500 drawn from the dictionary.⁴ Since nouns are most readily scaled for referential concreteness because they usually have the clearest referents, only homographs with two or more semantically distinct nominal referents were retained. The meanings of each homograph were scaled for primacy and concreteness by psychology students. One group of judges made the primacy evaluations by writing down in order "the first meaning that comes to mind when the word is seen [scored 3], then the second meaning that comes to mind, and the third last of all [scored 1]". Another group of judges rated these meanings on a 4-point scale for concreteness of referent, following Brown's (1958) definition: A concrete word (scored 3) has a clearer "thing" character, is more "picturable," has a more tangible or visual contour and a smaller denotation than an abstract word (scored 0), for example, *apple* is more concrete than *experience*.

On the basis of these judgments 24 homographs were selected for the inclusion in the interpretation test, 6 having a secondary-concrete meaning paired

⁴ An exhaustive list of noun homographs compiled by Joy Yeager and the first author, and copies of the interpretation and repertory tests, may be obtained from Norman F. Watt, William James 1260, 33 Kirkland Street, Cambridge, Massachusetts, 02138.

TABLE 1
MEAN PERCENTAGES OF REPERTORY ERRORS BY ALL GROUPS FOR
VARIOUS TYPES OF HOMOGRAPH MEANING

Repertory error type	No. items	Normals (<i>n</i> = 20)	Alcoholics (<i>n</i> = 22)	Schizophrenics (<i>n</i> = 44)	Aphasics (<i>n</i> = 11)	Groups combined (<i>n</i> = 97)
Secondary-concrete	6	36	44	48	48	45
Secondary-abstract	7	19	21	33	55	30
Primary-abstract	5	13	21	28	47	26
Primary-concrete	6	13	18	20	42	21
Total errors	24	20	26	33	48	31
All secondary	13	27	32	40	52	37
All primary	11	13	19	24	45	23
All concrete	12	25	31	34	45	33
All abstract	12	16	21	31	52	28

with a primary-abstract meaning, and 6 having a secondary-abstract meaning and a primary-concrete meaning. Three homographs each paired the other four possible combinations of primacy and concreteness of meaning. The meanings classified as primary had average primacy scores of at least 2.25, and their overall average (mean of means) was 2.7. The secondary meanings were rated no higher than 2.00 and their mean of means was 1.2. Thus, the primary meanings were very salient and the secondary meanings distinctly more obscure, typically being the *third* meaning recalled. The meanings classified as concrete had average concreteness ratings of at least 2.00, and their overall average was 2.7. The abstract meanings were rated 1.75 or less, and their overall average was .9.

The repertory test was designed to test knowledge of homograph meanings. As in Willner's (1965) format, each of the 24 homographs was followed by five nouns from which *S* had to select the one closest in meaning. A distractor alternative was included in 17 items. Each homograph had two potential meanings included in the interpretation test, but only half were assessed in the repertory test.

Scoring. There were 12 different types of error possible in the interpretation test, depending on the type of meaning required for correct interpretation and the type of meaning that was incorrect. For example, Error Type 1 consisted of selecting a primary-abstract meaning when the correct interpretation was secondary-concrete, and Error Type 2 was the reverse of this. A misinterpretation ratio was computed for each error type by dividing the number of wrong items by the number of wrong and right items combined [$W/(W + R)$]. An error score was thus expressed as a percentage of the items answered meaningfully (i.e., excluding the items where *S* chose the irrelevant alternative). Any *S* was dropped from the study if he chose the irrelevant meaning on more than half of the items for any of the 12 error types. Four schizophrenics were eliminated for this reason. This scoring procedure is different from the one used by Chapman et al. (1964), who subtracted the irrelevant choices from the wrong ones to arrive at the error score. They did not exclude any *S* for choosing too many irrelevant meanings.

RESULTS

Semantic Repertory

The results of the repertory test are presented in Table 1 and summarized by analysis of variance in Table 2. Through an oversight in constructing the repertory test, the number of items was not equal for all four types of homograph, so the errors were computed as a percentage of the items of each type in the test (5, 6, or 7, respectively). Table 1 gives the mean percentages for each group on each type. The analysis of variance in Table 2 was based on the raw scores after adjusting them for differences in the number of items present. The significant main effect for diagnosis shows that the groups differed in their total errors. The aphasics made more repertory errors than every other group, and the schizo-

TABLE 2
SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF
VARIANCE OF REPERTORY TEST ERRORS

Source of variation	SS	df	MS	F
Between Ss		96		
Diagnosis (D)	11998.34	3	3999.45	8.22***
Ss within groups				
(error _{DS})	45258.10	93	486.65	
Within Ss		291		
Primacy (P)	4456.14	1	4456.14	43.37***
D × P	370.93	3	123.64	1.20
P × Ss within groups				
(error _{DP})	9556.17	93	102.75	
Concreteness (C)	468.11	1	468.11	4.22**
D × C	1005.09	3	335.03	3.02*
C × Ss within groups				
(error _{DC})	10323.94	93	111.01	
P × C	1705.82	1	1705.82	20.73***
D × P × C	732.68	3	244.23	2.97*
PC × Ss within groups (error _{PC})	7654.06	93	82.30	

* $p < .10$.

** $p < .05$.

*** $p < .001$.

TABLE 3

TOTAL ERROR RATES ON THE INTERPRETATION TEST FOR ALL GROUPS
AS A FUNCTION OF REPERTORY ADEQUACY

Repertory item	Interpretation test error rate (%)				
	Normals (<i>n</i> = 20)	Alcoholics (<i>n</i> = 22)	Schizophrenics (<i>n</i> = 44)	Aphasics (<i>n</i> = 11)	Groups combined (<i>n</i> = 97)
Correct					
<i>M</i>	7.40	12.68	20.73	23.36	16.45
<i>SD</i>	5.16	15.35	13.62	11.78	13.88
Incorrect					
<i>M</i>	9.95	24.18	27.61	29.91	23.45
<i>SD</i>	15.63	24.22	20.88	21.16	21.96

phrenics more than the normals, but the alcoholics were not significantly different from either the normals or the schizophrenics. The significant main effect for primacy shows, as expected, that all groups knew more primary than secondary word meanings. The significant main effect for concreteness shows that more concrete than abstract repertory errors were made. This result is primarily due to the unusually high error rate for secondary-concrete meanings (cf. the strong Primacy \times Concreteness interaction), which is probably attributable to the presence of distractor alternatives with high "pull" for the secondary-concrete items. Data from an alternate form of the repertory test with different Ss showed the reverse order of secondary-concrete and secondary-abstract errors (a more logical result), which suggests that the error rate for secondary-concrete meanings in the present experiment may have been inflated artifactually.

None of the three interactions reflecting group comparisons of repertory error types reached statistical significance. From the insignificant Diagnosis \times Primacy interaction the authors conclude that neither the schizophrenics nor the aphasics were relatively deficient in knowledge of secondary word meanings. Specifically, the discrepancy between secondary and primary errors was no larger for the schizophrenics than the normals ($t = .58$) nor the alcoholics ($t = .76$) and was actually smaller for the aphasics than the normals ($t = -1.19$) and the alcoholics ($t = -1.05$). This result confirms Chapman's observation that schizophrenics and normals do not differ in their knowledge of secondary

meanings, rather than Willner's finding of the contrary. The result for the aphasics is consistent with Willner's finding for brain-damaged patients.

The trend toward significant Diagnosis \times Concreteness interaction is largely accounted for by the performance of the aphasics, whose discrepancy between abstract and concrete errors was larger than the normals' ($t = 2.08$, $p < .05$) and the alcoholics' ($t = 2.34$, $p < .05$) and slightly larger than the schizophrenics' ($t = 1.45$, $p < .20$). The schizophrenics showed a slight trend in the same direction when compared with the alcoholics ($t = 1.45$, $p < .20$) but not with the normals ($t = 1.09$, *ns*).

Relation of Repertory to Interpretation in Context

It will be recalled that the repertory test included half of the meanings used in the interpretation test, so it was possible to examine the role of repertory adequacy in determining errors of interpretation for those 24 items. The results with reference to the total error rates on the interpretation test (Table 3) were analyzed by unweighted-means analysis of variance. There was a significant main effect for diagnosis ($F = 6.37$, $p < .001$), showing that the groups differed in their overall error rates on the interpretation test, and a significant main effect for repertory adequacy ($F = 7.63$, $p < .01$), indicating that all groups combined made more errors of interpretation when the meaning required for correct interpretation was not known. The in-

^a All significance levels were based on two-tailed tests, unless otherwise indicated.

teraction of Diagnosis \times Repertory Adequacy did not approach significance ($F = .54$), indicating there were no differences among the groups in the proportion of interpretation errors due to unknown meanings.

Table 4 presents the interpretation error rates for all groups combined, broken down by the four varieties of homograph meaning required and by repertory adequacy. When the homograph meaning was correct, the overall error rates for the four types of meaning did not differ significantly. But the results in the lower half of the table show clearly that inadequate vocabulary accounts for more errors of interpretation when the correct interpretation requires a secondary-abstract meaning than in any other type. The low secondary-concrete error rate there corroborates the authors' earlier impression that some secondary-concrete repertory errors were due to artifacts in the repertory test rather than true ignorance of the meaning. When the data were broken down further by diagnostic groups, some of the error frequencies were too small for meaningful analysis, but it was clear from examination of the data that there were no further differences among the groups.

Semantic Interpretation in Context

Table 5 summarizes the overall results for primacy bias, concreteness bias, and total errors of interpretation in context. The groups differed significantly in total errors, with the

TABLE 4
PERCENTAGE OF INTERPRETATION TEST ERRORS BY ALL GROUPS COMBINED FOR EACH TYPE OF HOMOGRAPH MEANING REQUIRED AS A FUNCTION OF REPERTORY ADEQUACY

Repertory Item	Interpretation test error rate (%)		
	Type of meaning required		Total
	Concrete	Abstract	
Correct			
Primary	17	10	14
Secondary	10	17	14
Total	14	14	14
Incorrect			
Primary	29	28	29
Secondary	17	45	29
Total	21	39	29

alcoholics making more errors than the normals ($t = 2.25$, $p < .05$), the schizophrenics more than the alcoholics ($t = 2.39$, $p < .02$), and the aphasics more than the schizophrenics ($t = 1.74$, $p < .10$). Analysis of variance of the primary and secondary errors showed a significant main effect for diagnosis ($F = 8.93$, $p < .001$) but only slight evidence of a primacy effect ($F = 1.66$, $p = .20$) and no Primacy \times Diagnosis interaction ($F = .61$, ns). Comparing specifically the primacy biases (primary-secondary errors), the schizophrenics slightly exceeded the normals ($t = 1.61$, $p < .20$) and the alcoholics ($t = 1.24$,

TABLE 5
SUMMARY OF ERROR RATES ON THE INTERPRETATION TEST

Type of Error	No. items	Mean interpretation error rate (%)			
		Normals ($n = 20$)	Alcoholics ($n = 22$)	Schizophrenics ($n = 44$)	Aphasics ($n = 11$)
Primary errors (Error Types 1, 3, 7, & 9)	18	10.10	13.74	21.81	26.63
Secondary errors (Error Types 2, 4, 8, & 10)	18	9.70	14.21	16.75	23.78
Primacy bias (Primary-secondary errors)		0.40	-0.47	5.06*	2.85
Concrete errors (Error Types 2, 3, 6, & 12)	18	13.45	19.18	24.86	32.18
Abstract errors (Error Types 1, 4, 5, & 11)	18	5.50	10.64	19.46	23.09
Concreteness bias (Concrete-abstract errors)		7.95	8.54	5.40	9.09
Total errors	48	8.95	14.05	20.36	26.82

* The primacy bias of the schizophrenics showed a slight trend toward significant difference from the normals ($p < .20$) and the alcoholics ($p < .20$).

$p < .20$) but not the aphasics. Only if the data for the normals and alcoholics are combined, and the comparison is given the advantage of a one-tailed test of significance, do the data support Chapman's theory of greater primacy bias in schizophrenics ($t = 1.69$, $df = 84$, $p < .05$, one-tailed test).

The analysis of variance of the concrete and abstract errors found a significant main effect for diagnosis ($F = 11.67$, $p < .001$) and for concreteness ($F = 22.31$, $p < .001$) but no Concreteness \times Diagnosis interaction ($F = .24$, ns). None of the group comparisons of concreteness bias approached statistical significance.

Examination of the results for Error Types 1-4, which included half of the 48 test items, provided an additional test of the interaction Primacy Bias \times Concreteness Bias. Again the authors found a significant main effect for diagnosis and for concreteness, and a significant main effect for primacy, which corroborated for the first time Chapman's hypothesis of a general human inclination to make more primary than secondary errors. However, there was no interaction of primacy bias and concreteness bias for the groups combined, and none of the interactions involving diagnosis approached significance.

Correlations of Semantic Measures with Other Variables

The correlations of the principal language measures with verbal intelligence, education, and length of hospitalization were analyzed for all groups but the aphasics. The most striking finding was that the language measures of the schizophrenics and alcoholics correlated higher with verbal intelligence than with education in almost every case, whereas the performance of the normals consistently correlated as highly with education as with verbal intelligence, especially on the measures of interpretation in context. Thus, when the variance for education was partialled out, the correlation between total interpretational errors and verbal intelligence dropped from $-.49$ ($p < .05$) to $-.12$ (ns) for the normals, but remained virtually unchanged for the schizophrenics ($-.75$ to $-.73$) and alcoholics ($-.63$ to $-.61$), all significant at .01. The analysis for total repertory errors was simi-

lar, though less extreme. Therefore, it seems clear that intellectual training plays a more important role in determining these language functions for normals than for either of the patient groups. Considering that the measure of verbal intelligence at the same time assesses attentiveness or presence of mind, it could be inferred that semantic functioning in mental patients is dependent to a great extent on current attention to the task. If this is true, then it is remarkable that the phenomenon is as marked in the alcoholics as the schizophrenics, which would imply that it is not specific to schizophrenia but is general in chronic patients, perhaps reflecting the effects of drugs, long hospitalization, or variability in the motivation to comply with the requirements of the testing situation.

The inference about hospitalization effects is challenged somewhat by the fact that only 1 of the 28 correlations with length of hospitalization was significant. That correlation showed that schizophrenics who had been hospitalized longer had *less* primacy bias, mainly because they made more secondary errors of interpretation in context. This result seems to be incongruous with Chapman's theory, unless it is assumed that the primacy bias diminishes with increasing chronicity, which seems unlikely.

SUMMARY AND DISCUSSION

Groups combined. All groups knew more primary than secondary word meanings; that is, meanings which come readily to mind are known by more people than less salient meanings. In short, "retrieval readiness" is positively related to comprehension. The Ss in this experiment knew more abstract than concrete meanings, but it is believed that this was principally due to a few strong distractors among the secondary-concrete items. (It would have served the purposes of this study better to have no distractors at all on the repertory test, but that would deviate from Willner's procedure.) Comparison of the primary-concrete versus primary-abstract items, and results from a separate experiment, strongly indicate that if artifacts are absent, more concrete than abstract meanings are known. More errors of interpretation were made when the homograph meaning was not

known than when it was known, especially when the correct interpretation was secondary and abstract.

Regarding interpretation of homographs in context, there was a strong concreteness bias but less clear evidence of an overall primacy bias. Further analysis revealed an inverse relation between strength of the primacy bias and degree of ambiguity in the sentence. (Sentence ambiguity, or general difficulty in correctly interpreting the item, was rated intuitively by the second author; it usually boiled down to the number of misleading cues in the sentence.) There was a clear primacy bias (+ 4.5%) in sentences with low ambiguity, but an even stronger *secondary* bias (- 6.9%) in sentences with high ambiguity. It appears that there is a general human tendency to err by choosing a more common meaning when the interpretation is relatively uncomplicated by misleading cues, as if such errors were automatic ("by force of habit"). However, when the sentence is highly ambiguous, Ss err by choosing the less common meaning, as if they thought, "an interpretation that requires reflection must imply an obscure meaning." This account of the primacy bias is also consistent with Chapman's finding a marked primacy bias on an instrument intentionally constructed with sentences of low ambiguity. In the present experiment there was no interaction between primacy bias and concreteness bias.

Group differences. Mental patients made more total errors on the raw experimental measures than the normals, and the magnitude of semantic "impairment" was related to severity of language disturbance clinically. Consistently the aphasics made the most errors, followed by the schizophrenics, the alcoholics and the normals, in that order. The same order obtained for percentage of irrelevant alternatives chosen on the interpretation test, the rates being 6% (aphasics), 2% (schizophrenics), 1% (alcoholics), and .3% (normals). This implies that a general factor of inattention to the task (or inability to focus attention meaningfully) accounts for a substantial part of the errors in repertory and interpretation. The fact that the alcoholics differed from the normals on most of the measures indicates that semantic functioning

is influenced by chronic hospitalization or some clinical aspect of alcoholism.

With regard to the principal hypotheses, the aphasics knew fewer abstract meanings than the control groups, which confirms previous research and the general clinical impression that abstract vocabulary is most affected in aphasia. This repertory deficiency did not handicap them, especially on the interpretation test. A deficiency of abstract meanings was *not* found for the schizophrenics, nor any evidence of a stronger bias toward concrete errors of interpretation in context among the mental patients. Thus, except for the loss of abstract meanings in the aphasics, the principle of linguistic regression found no support in this study.

The schizophrenics showed a somewhat stronger primacy bias than the control groups in interpreting words in context. This supports Chapman's (1964) theory, but not very strongly. It should be noted that, although the homographs were selected by clearly more stringent criteria for primacy than Chapman's, the sentences spanned a much wider range of semantic ambiguity than his. The authors found that Chapman's theory held for relatively unambiguous items, all groups showing a primacy bias and the schizophrenics more than the others. However, finding a *secondary* bias on highly ambiguous items and no group differences suggests that his theory may not hold for all types of disambiguation measures.

The present study confirmed Chapman's finding that schizophrenics do not "lose" secondary meanings from their vocabulary. Willner's (1965) contrary finding that schizophrenics knew fewer unusual (secondary) meanings may be due to his test being more difficult. Normals in the present study made 27% errors on the secondary items, whereas Willner's normals made 45-50% errors on his Unusual Meanings Vocabulary Test. Further, Willner⁶ does not "really believe that the unusual meanings of words are permanently 'lost,' but rather that schizophrenic patients have significantly more difficulty in retrieving these meanings." This implies that schizophrenics might perform poorly on Willner's

⁶ A. Willner, personal communication, 1968.

test (in which correct answers require weaker mediating responses) for the same reason that they might show a strong primacy bias when interpreting words in context (overreliance on strongest mediating responses). Such an explanation would reconcile Willner's seemingly different results with those in this study and in Chapman's.

REFERENCES

- BABCOCK, H. An experiment in the measurement of mental deterioration. *Archives of Psychology*, 1930, 18, No. 117.
- BLEULER, E. *Dementia praecox or the group of schizophrenias*. New York: International Universities Press, 1950.
- BROWN, R. *Words and things*. New York: Free Press, 1958.
- CHAPMAN, L. J., CHAPMAN, J. P., & MILLER, G. A. A theory of verbal behavior in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1964.
- DOMARUS, E. VON. The specific laws of logic in schizophrenia. In J. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: University of California Press, 1944.
- FEFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genetic Psychology Monographs*, 1949, 39, 151-204.
- FLAVELL, J. Abstract thinking and social behavior in schizophrenia. *Journal of Abnormal and Social Psychology*, 1956, 52, 208-211.
- GOLDSTEIN, K. *Language and language disturbances: Aphasic symptom complexes and their significance for medicine and theory of language*. New York: Grune & Stratton, 1948.
- GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychological Monographs*, 1941, 53(2, Whole No. 239).
- HARRINGTON, R., & EHRLMANN, J. C. Complexity of response as a factor in the vocabulary performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1954, 49, 362-364.
- LUBIN, A., GIESEKING, C. F., & WILLIAMS, H. L. Direct measurement of cognitive deficit in schizophrenia. *Journal of Consulting Psychology*, 1962, 26, 139-143.
- MILGRAM, N. Preference for abstract versus concrete word meanings in schizophrenic and brain-damaged patients. *Journal of Clinical Psychology*, 1959, 15, 207-212.
- PIRO, S. Contributo allo studio della dissociazione semantica nel linguaggio schizofrenico. *Acta neurol. Napoli*, 1958, 13, 392-518.
- TERMAN, L. M. The vocabulary test as a measure of intelligence. *Journal of Educational Psychology*, 1918, 9, 452-466.
- VIGOTSKY, L. S. Thought in schizophrenia. *Archives of Neurological Psychiatry*, 1934, 31, 1063-1077.
- WERNER, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.
- WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. F. Direct measurement of cognitive deficit in brain-injured patients. *Journal of Consulting Psychology*, 1959, 23, 300-305.
- WILLNER, A. Impairment of knowledge of unusual meanings of familiar words in brain damage and schizophrenia. *Journal of Abnormal Psychology*, 1965, 70, 405-411.
- YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychological Review*, 1941, 48, 261-267.

(Received October 3, 1968)